BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C11N 5/10, C07K 14/705, 16/28 — A2

(11) International Publication Number: **WO 95/23859**

(43) International Publication Date: 8 September 1995 (08.09.95)

(21) International Application Number: PCT/US95/02576

(22) International Filing Date: 2 March 1995 (02.03.95)

(30) Priority Data:
08/205,697 2 March 1994 (02.03.94) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/205,697 (CIP)
Filed on 2 March 1994 (02.03.94)

(71) Applicants *(for all designated States except US)*: BRIGHAM AND WOMEN'S HOSPITAL [US/US]; 75 Francis Street, Boston, MA 02115 (US). DANA-FARBER CANCER INSTITUTE [US/US]; 44 Binney Street, Boston, MA 02115 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SHARPE, Arlene, H. [US/US]; 305 Walnut Street, Brookline, MA 02146 (US). BORRIELLO, Francescopaolo [US/US]; Apartment #3, 20 Perry Street, Brookline, MA 02146 (US). FREEMAN, Gordon, J. [US/US]; 305 Walnut Street, Brookline, MA 02146 (US). NADLER, Lee, M. [US/US]; 36 Cross Hill Road, Newton, MA 02159 (US).

(74) Agents: MANDRAGOURAS, Amy, E. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NOVEL FORMS OF T CELLS COSTIMULATORY MOLECULES AND USES THEREFOR

(57) **Abstract**

Novel structural forms of T cell costimulatory molecules are described. These structural forms comprise a novel structural domain or have a structural domain deleted or added. The structural forms correspond to naturally-occuring alternatively spliced forms of T cell costimulatory molecules or variants thereof which can be produced by standard recombinant DNA techniques. In one embodiment, the T cell costimulatory molecule of the invention contains a novel cytoplasmic domain. In another embodiment, the T cell costimulatory molecule of the invention contains a novel signal peptide domain or has an immunoglobulin variable region-like domain deleted. The novel structural forms of T cell costimulatory molecules can be used to identify agents which stimulate the expression of alternative forms of costimulatory molecules and to identify components of the signal transduction pathway which results in costimulation of T cells.

NOVEL FORMS OF T CELL COSTIMULATORY MOLECULES AND USES THEREFOR

Background of the Invention

For CD4+ T lymphocyte activation to occur, two distinct signals must be delivered by antigen presenting cells to resting T lymphocytes (Schwartz, R.H. (1990) *Science* 248:1349-1356; Williams, I.R. and Unanue, E.R. (1991) *J. Immunol.* 147:3752-3760; Mueller, D.L. et al., (1989) *J. Immunol.* 142:2617-2628). The first, or primary, activation signal is mediated physiologically by the interaction of the T cell receptor/CD3 complex (TcR/CD3) with MHC class II-associated antigenic peptide and gives specificity to the immune response. The second signal, the costimulatory signal, regulates the T cell proliferative response and induction of effector functions. Costimulatory signals appear pivotal in determining the functional outcome of T cell activation since delivery of an antigen-specific signal to a T cell in the absence of a costimulatory signal results in functional inactivation of mature T cells, leading to a state of tolerance (Schwartz, R.H. (1990) *Science* 248:1349-1356).

Molecules present on the surface of antigen presenting cells which are involved in T cell costimulation have been identified. These T cell costimulatory molecules include murine B7-1 (mB7-1; Freeman, G.J. et al., (1991) *J. Exp. Med.* 174:625-631), and the more recently identified murine B7-2 (mB7-2; Freeman, G.J. et al., (1993) *J. Exp. Med.* 178:2185-2192). Human counterparts to the murine B7-1 and B7-2 molecules have also been described (human B7-1 (hB7-1) Freedman, A.S. et al., (1987) *J. Immunol.* 137:3260-3267; Freeman, G.J. et al., (1989) *J. Immunol.* 143:2714-2722; and human B7-2 (hB7-2); Freeman, G.J. et al., (1993) *Science* 262:909-911; Azuma, M. et al. (1993) *Nature* 366:76-79). The B7-1 and B7-2 genes are members of the immmunoglobulin gene superfamily.

B7-1 and B7-2 display a restricted pattern of cellular expression, which correlates with accessory cell potency in providing costimulation (Reiser, H. et al. (1992; *Proc. Natl. Acad. Sci. USA* 89:271-275; Razi-Wolf Z. et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:4210-4214; Galvin, F. et al. (1992) *J. Immunol.* 149:3802-3808; Freeman, G.J. et al., (1993) *J. Exp. Med.* 178:2185-2192). For example, B7-1 has been observed to be expressed on activated B cells, T cells and monocytes but not on resting B cells, T cells or monocytes, and its expression can be regulated by different extracellular stimuli (Linsley, P.S. et al., (1990) *Proc. Natl.Acad. Sci. USA* 87:5031-5035; Linsley, P.S. et al., (1991) *J. Exp. Med.* 174:561-569; Reiser, H. et al. (1992); *Proc. Natl. Acad. Sci. USA* 89:271-275; Gimmi, C.D. et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:6575-6579; Koulova, L. et al. (1991) *J. Exp. Med.* 173:759-762; Azuma, M. et al. (1993) *J. Exp. Med.* 177:845-850; Sansom, D.M. et al. (1993) *Eur. J. Immunol.* 23:295-298)

Both B7-1 and B7-2 are counter-receptors for two ligands, CD28 and CTLA4, expressed on T lymphocytes (Linsley, P.S. et al., (1990) *Proc. Natl.Acad. Sci. USA* 87:5031-

different cytoplasmic domains. The existence of alternative cytoplasmic domain forms of T cell costimulatory molecules supports a functional role for the cytoplasmic domain in transmitting an intracellular signal within a cell which expresses the costimulatory molecule on its surface. This indicates that costimulatory molecules not only trigger an intracellular signal in T cells, but may also deliver a signal to the cell which expresses the costimulatory molecule. This is the first evidence that the interaction between a costimulatory molecule on one cell and its receptor on a T cell may involve bidirectional signal transduction between the cells (rather than only unidirectional signal transduction to the T cell).

In yet another aspect of the invention, proteins that bind CD28 and/or CTLA4 and contain a novel signal peptide domain are provided. T cell costimulatory molecule genes which contain exons encoding different signal peptide domains which are used in an alternate manner have been discovered. Alternative splicing of mRNA transcripts of the gene can generate native T cell costimulatory molecules with different signal peptide domains. The existence of alternative signal peptide domain forms of T cell costimulatory molecules also suggests a functional role for the signal peptide of T cell costimulatory molecules.

Still another aspect of the invention pertains to isolated proteins that bind CD28 and/or CTLA4 in which a structural domain has been deleted or added, and isolated nucleic acids encoding such proteins. In a preferred embodiment, the protein (e.g., B7-1) has an immunoglobulin constant-like domain deleted (i.e., an immunoglobulin variable-like domain is linked directly to a transmembrane domain). In another embodiment, the protein has an immunoglobulin variable-like domain deleted (i.e., a signal peptide domain is linked directly to an immunoglobulin constant-like domain).

An isolated nucleic acid molecule of the invention can be incorporated into a recombinant expression vector and transfected into a host cell to express a novel structural form of a T cell costimulatory molecule. The isolated nucleic acids of the invention can further be used to create transgenic and homologous recombinant non-human animals. The novel T cell costimulatory molecules provided by the invention can be used to trigger a costimulatory signal in a T lymphocyte. These molecules can further be used to raise antibodies against novel structural domains of costimulatory molecules. The novel T cell costimulatory molecules of the invention can also be used to identify agents which stimulate the expression of alternative forms of costimulatory molecules and to identify components of the signal transduction pathway induced in a cell expressing a costimulatory molecule in response to an interaction between the costimulatory molecule and its receptor on a T lymphocyte.

Brief Description of the Drawings

Figure 1 is a photograph of an agarose gel depicting the presence of mB7-1 cytoplasmic domain II-encoding exon 6 in mB7-1 cDNA, determined by nested Reverse Transcriptase Polymerase Chain Reaction (RT-PCR).

disclosed forms of T cell costimulatory molecules, e.g., forms which result from alternative splicing of a primary mRNA transcript of a gene encoding a T cell costimulatory molecule.

Accordingly, one aspect of the invention relates to isolated nucleic acids encoding T cell costimulatory molecules corresponding to naturally-occurring alternatively spliced forms or variants thereof, and uses therefor. Another aspect of the invention pertains to novel structural forms of T cell costimulatory molecules which are produced by transcription and translation of the nucleic acid molecules of the invention, and uses therefor. This invention further pertains to isolated nucleic acids encoding novel structural domains of T cell costimulatory molecules, isolated polypeptides encoded therein, and uses therefor.

The various aspects of this invention are described in detail in the following subsections. Forming part of the present disclosure is the appended Sequence Listing. The numerous nucleotide and amino acid sequences presented in the Sequence Listing are summarized below.

SEQ ID NO: 1 - nucleotide sequence of mouse B7-1 exons 1-2-3-4-6
SEQ ID NO: 2 - amino acid sequence of mouse B7-1 protein encoded by exons 1-2-3-4-6
SEQ ID NO: 3 - nucleotide sequence of mouse B7-1 exons 1-2-3-4-5-6
SEQ ID NO: 4 - nucleotide sequence of mouse B7-1 exon 6 (CytII)
SEQ ID NO: 5 - amino acid sequence of mouse B7-1 peptide encoded by exon 6 (CytII)
SEQ ID NO: 6 - nucleotide sequence of mouse B7-1 full-length exon 1
SEQ ID NO: 7 - nucleotide sequence of mouse B7-1 promoter
SEQ ID NO: 8 - nucleotide sequence of B7-1 exons 1-3-4-5
SEQ ID NO: 9 - amino acid sequence of mB7-1 protein encoded by exons 1-3-4-5
SEQ ID NO: 10 - nucleotide sequence of mouse B7-1 exons 1-3-4-6
SEQ ID NO: 11 - amino acid sequence of mouse B7-1 protein encoded by exons 1-3-4-6
SEQ ID NO: 12 - nucleotide sequence of mouse B7-2 exons m1B-2-3-4-5
SEQ ID NO: 13 -amino acid sequence of mouse B7-2 protein encoded by exons m1B-2-3-4-5
SEQ ID NO: 14 - nucleotide sequence of mouse B7-2 exon m1B
SEQ ID NO: 15 - amino acid sequence of mouse B7-2 peptide encoded by exon m1B
SEQ ID NO: 16 - nucleotide sequence of mouse B7-1 exons 1-2-3-4-5 (as disclosed in Freeman, G. J. et al. (1991) *J. Exp. Med.* 174:625-631)
SEQ ID NO: 17 - amino acid sequence of mouse B7-1 protein encoded by exons 1-2-3-4-5
SEQ ID NO: 18 - nucleotide sequence of human B7-1 exons 1-2-3-4-5 (as disclosed in Freeman, G.J. et al. (1989) *J. Immunol.* 143:2714-2722)
SEQ ID NO: 19 - amino acid sequence of human B7-1 protein encoded by exons 1-2-3-4-5
SEQ ID NO: 20 - nucleotide sequence of mouse B7-2 exons m1A-2-3-4-5 (as disclosed in Freeman, G.J. et al. (1993) *J. Exp. Med.* 178:2185-2192)
SEQ ID NO: 21 -amino acid sequence of mouse B7-2 protein encoded by exons m1A-2-3-4-5

and RNA and can be either double stranded or single stranded. Preferably, the isolated nucleic acid molecule is a cDNA.

A. Nucleic Acids Encoding Novel Cytoplasmic Domains

One aspect of the invention pertains to isolated nucleic acids that encode T cell costimulatory molecules, each containing a novel cytoplasmic domain. It has been discovered that a gene encoding a costimulatory molecule can contain multiple exons encoding different cytoplasmic domains. In addition, naturally-occurring mRNA transcripts have been discovered which encode different cytoplasmic domain forms of T cell costimulatory molecules. Thus, one embodiment of the invention provides an isolated nucleic acid encoding a protein which binds CD28 or CTLA4 and comprises a contiguous nucleotide sequence derived from at least one T cell costimulatory molecule gene. In this embodiment, the nucleotide sequence can be represented by a formula A-B-C-D-E, wherein

A comprises a nucleotide sequence of at least one first exon encoding a signal peptide domain,

B comprises a nucleotide sequence of at least one second exon of a T cell costimulatory molecule gene, wherein the at least one second exon encodes an immunoglobulin variable region-like domain,

C comprises a nucleotide sequence of at least one third exon of a T cell costimulatory molecule gene, wherein the at least one third exon encodes an immunoglobulin constant region-like domain,

D comprises a nucleotide sequence of at least one fourth exon of a T cell costimulatory molecule gene, wherein the at least one fourth exon encodes a transmembrane domain, and

E comprises a nucleotide sequence of at least one fifth exon of a T cell costimulatory molecule gene, wherein the at least one fifth exon encodes a cytoplasmic domain,

with the proviso that E does not comprise a nucleotide sequence encoding a cytoplasmic domain selected from the group consisting of SEQ ID NO:25 (mB7-1), SEQ ID NO:27 (hB7-1), SEQ ID NO:29 (mB7-2) and SEQ ID NO:31 (hB7-2).

In the formula, A, B, C, D, and E are contiguous nucleotide sequences linked by phosphodiester bonds in a 5' to 3' orientation from A to E. According to the formula, A can be a nucleotide sequence of an exon which encodes a signal peptide domain of a heterologous protein which efficiently expresses transmembrane or secreted proteins, such as the oncostatin M signal peptide. Preferably, A comprises a nucleotide sequence of at least one exon which encodes a signal peptide domain of a T cell costimulatory molecule gene. It is

cytoplasmic domain of the T cell costimulatory molecule. Additionally, a second, alternative cytoplasmic domain of another T cell costimulatory molecule is likely to be homologous to the Cyt II domain of mB7-1. For example, the first cytoplasmic domains of mB7-1, hB7-1, mB7-2 and hB7-2 display between 4 % and 26 % amino acid identity (see Freeman, G.J. et al. (1993) *J. Exp. Med.* 178:2185-2192). Accordingly, in one embodiment, an alternative cytoplasmic domain of a T cell costimulatory molecule has an amino acid sequence that is at least about 5 % to 25 % identical in sequence with the amino acid sequence of mB7-1 Cyt II (shown in SEQ ID NO: 5).

Another embodiment of the invention provides an isolated nucleic acid encoding a protein which binds CD28 or CTLA4 and is encoded by a T cell costimulatory molecule gene having at least one first exon encoding a first cytoplasmic domain and at least one second exon encoding a second cytoplasmic domain. The at least one first cytoplasmic domain exon of the gene comprises a nucleotide sequence selected from the group consisting of a nucleotide sequence of SEQ ID NO:25 (mB7-1), SEQ ID NO:27 (hB7-1), SEQ ID NO:29 (mB7-2) and SEQ ID NO:31 (hB7-2). In this embodiment, the isolated nucleic acid includes a nucleotide sequence encoding at least one second cytoplasmic domain. Preferably, the isolated nucleic acid does not comprise a nucleotide sequence encoding a first cytoplasmic domain (i.e., the nucleic acid comprises an alternative splice form of a transcript of the gene in which the exon encoding the first cytoplasmic domain, e.g., exon 5, has been excised from the transcript). Preferred T cell costimulatory molecule genes from which nucleotide sequences can be derived include B7-1 and B7-2.

In yet another embodiment, the isolated nucleic acid of the invention encodes a protein which binds CD28 or CTLA4 and comprises a nucleotide sequence shown in SEQ ID NO: 1. This nucleotide sequence corresponds to a naturally-occurring alternatively spliced form of mB7-1 which includes the nucleotide sequences of exons 1-2-3-4-6. Alternatively, the isolated nucleic acid comprises a nucleotide sequence shown in SEQ ID NO: 3, which corresponds to a naturally-occurring alternatively spliced form of mB7-1 comprising the nucleotide sequences of exons 1-2-3-4-5-6.

B. Nucleic Acids Encoding Novel Signal Peptide Domains

Other aspects of this invention pertain to isolated nucleic acids which encode T cell costimulatory molecules containing novel signal peptide domains. It has been discovered that a gene encoding a costimulatory molecule can contain multiple exons encoding different signal peptide domains and that mRNA transcripts occur in nature which encode different signal peptide domain forms of T cell costimulatory molecules. Thus, isolated nucleic acids which encode proteins which bind CD28 or CTLA4 and comprise contiguous nucleotide sequences derived from at least one T cell costimulatory molecule gene are within the scope of this invention. The nucleotide sequence can be represented by a formula A-B-C-D-E, wherein

In yet another embodiment of the invention, the isolated nucleic acid encodes a protein which binds CD28 or CTLA4 and is encoded by a T cell costimulatory molecule gene having at least one first exon encoding a first signal peptide domain and at least one second exon encoding a second signal peptide domain. The at least one first exon comprises a nucleotide sequence selected from the group consisting of a nucleotide sequence of SEQ ID NO:33 (mB7-1), SEQ ID NO:35 (hB7-1), SEQ ID NO:37 (mB7-2) and SEQ ID NO:39 (hB7-2) and SEQ ID NO:41 (hB7-2). In this embodiment, the isolated nucleic acid includes a nucleotide sequence encoding at least one second signal peptide domain. Preferably, the isolated nucleic acid does not comprise a nucleotide sequence encoding the first signal peptide domain (i.e., the nucleic acid comprises an alternative splice form of a transcript of the gene in which the exon encoding a first signal domain has been excised from the transcript). Preferred T cell costimulatory molecule gene from which nucleotide sequences can be derived include B7-1 and B7-2.

C. Nucleic Acids Encoding Proteins With Domains Deleted or Added

Another aspect of the invention pertains to isolated nucleic acids encoding T cell costimulatory molecules having structural domains which have been deleted or added. This aspect of the invention is based, at least in part, on the discovery that alternative splicing of mRNA transcripts encoding T cell costimulatory molecules generates transcripts in which an exon encoding a structural domain has been excised or in which at least two exons encoding two forms of a structural domain are linked in tandem. In one embodiment, the nucleic acid is one in which an exon encoding an IgV-like domain has been deleted (i.e., the signal peptide domain exon is linked directly to the IgC-like domain exon). Accordingly, in one embodiment, the isolated nucleic acid encodes a protein comprising a contiguous nucleotide sequence derived from at least one T cell costimulatory molecule gene, the nucleotide sequence represented by a formula A-B-C-D, wherein

A comprises a nucleotide sequence of at least one first exon of a T cell costimulatory molecule gene, wherein the at least one first exon encodes a signal peptide domain,

B comprises a nucleotide sequence of at least one second exon of a T cell costimulatory molecule gene, wherein the at least one second exon encodes an immunoglobulin constant region-like domain,

C comprises a nucleotide sequence of at least one third exon of a T cell costimulatory molecule gene, wherein the at least one third exon encodes a transmembrane domain, and

D comprises a nucleotide sequence of at least one fourth exon of a T cell costimulatory molecule gene, wherein the at least one fourth exon encodes a cytoplasmic domain.

cytoplasmic domain of mB7-1. The amino acid sequence of the protein encoded by this nucleic acid is shown in SEQ ID NO: 65. Naturally-occurring mRNA transcripts encoding murine B7-1 have been detected in which the exon encoding the IgC-like domain (i.e, exon 3) has been excised and the exon encoding the IgV-like domain (i.e., exon 2) is spliced to the exon encoding the transmembrane domain (i.e., exon 4) (see Example 7). When expressed in a host cell, the IgV-like isoform of mB7-1 is capable of binding to both mouse CTLA4 and mouse CD28 and can trigger a costimulatory signal in a T cell such that the T cell proliferates and produces interleukin-2 (see Example 7).

Yet another aspect of this invention features an isolated nucleic acid encoding a T cell costimulatory molecule which contains exons in addition to a known or previously identified form of the T cell costimulatory molecule. For example, a naturally-occurring murine B7-1 mRNA transcript has been identified which contains two cytoplasmic domain-encoding exons in tandem, i.e., the transcript contains exons 1-2-3-4-5-6 (the nucleotide sequence of which is shown in SEQ ID NO: 3). Since there is an in-frame termination codon within exon 5, translation of this transcript produces a protein which contains only the Cyt I cytoplasmic domain. However, if desired, this termination codon can be mutated by standard site-directed mutagenesis techniques to create a nucleotide sequence which encodes an mB7-1 protein containing both a Cyt I and a Cyt II domain in tandem.

II. Isolation of Nucleic Acids of the Invention

An isolated nucleic acid having a nucleotide sequence disclosed herein can be obtained by standard molecular biology techniques. For example, oligonucleotide primers suitable for use in the polymerase chain reaction (PCR) can be prepared based upon the nucleotide sequences disclosed herein and the nucleic acid molecule can be amplified from cDNA and isolated. At least one oligonucleotide primer should be complimentary to a nucleotide sequence encoding an alternative structural domain. It is even more preferable that at least one oligonucleotide primer span a novel exon junction created by alternative splicing. For example, an oligonucleotide primer which spans the junction of exon 4 and exon 6 can be used to preferentially amplify a murine B7-1 cDNA that contains the second, alternative cytoplasmic domain (e.g., a cDNA which contains exons 1-2-3-4-6; SEQ ID NO: 1). Alternatively, an oligonucleotide primer complimentary to a nucleotide sequence encoding a novel alternative structural domain can be used to screen a cDNA library to isolate a nucleic acid of the invention.

Isolated nucleic acid molecules having nucleotide sequences other than those specifically disclosed herein are also encompassed by the invention. For example, novel structural forms of B7-1 from species other than mouse are within the scope of the invention (e.g., alternatively spliced forms of human B7-1). Likewise, novel structural forms of B7-2 from species other than mouse are also within the scope of the invention (e.g., alternatively spliced forms of human B7-2). Furthermore, additional alternatively spliced forms for

or a portion of a nucleotide sequence encoding the costimulatory molecule (e.g., having all or a portion of a nucleotide sequence shown in SEQ ID NO: 16, 18, 20, 22 and 24). For costimulatory molecules whose genes have been mapped to a particular chromosome, a chromosome-specific library rather than a total genomic DNA library can be used. For example, hB7-1 has been mapped to human chromosome 3 (see Freeman, G.J. et al. (1992) *Blood* 79:489-494; and Selvakumar, A. et al. (1992) *Immunogenetics* 36:175-181. Genomic clones can be sequenced by conventional techniques and novel exons identified. A probe corresponding to a novel exon can then be used to detect the nucleotide sequence of this exon in mRNA transcripts encoding the costimulatory molecule (e.g., by screening a cDNA library or by PCR).

A more preferred approach for identifying and isolating nucleic acid encoding a novel structural domain of a T cell costimulatory molecule is by "exon trapping". Exon trapping is a technique that has been used successfully to identify and isolate novel exons (see e.g. Duyk, G.M. et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:8995-8999; Auch, D. and Reth, M. (1990) *Nucleic Acids Res.* 18:6743-6744; Hamaguchi, M. et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:9779-9783; and Krizman, D.B and Berget, S.M. (1993) *Nucleic Acids Res.* 21:5198-5202). The approach of exon trapping can be applied to the isolation of exons encoding novel structural domains of T cell costimulatory molecules, such as a novel alternative cytoplasmic domain of human B7-1, as described in Example 5.

In addition to the isolated nucleic acids encoding naturally-occurring alternatively spliced forms of T cell costimulatory molecules provided by the invention, it will be appreciated by those skilled in the art that nucleic acids encoding variant alternative forms, which may or may not occur naturally, can be obtained used standard recombinant DNA techniques. The term "variant alternative forms" is intended to include novel combinations of exon sequences which can be created using recombinant DNA techniques. That is, novel exons encoding structural domains of T cell costimulatory molecules, either provided by the invention or identified according to the teachings of the invention, can be "spliced", using standard recombinant DNA techniques, to other exons encoding other structural domains of the costimulatory molecule, regardless of whether the particular combination of exons has been observed in nature. Thus, novel combinations of exons can be linked *in vitro* to create variant alternative forms of T cell costimulatory molecules. For example, the structural form of murine B7-1 which has the signal peptide domain directly joined to the IgC-like domain (ie., which has the IgV-like domain deleted) has been observed in nature in combination with the cytoplasmic domain encoded by exon 5. However, using conventional techniques, an alternative structural form can be created in which the IgV-like domain is deleted and the alternative cytoplasmic domain is encoded by exon 6. In another example, a murine B7-1 cDNA containing exons 1-2-3-4-5-6 can be mutated by site-directed mutagenesis to change a stop codon in exon 5 to an amino acid encoding-codon such that an mB7-1 protein can be produced which contains both a Cyt I domain and a Cyt II domain in tandem. Additionally,

forms of T cell costimulatory molecules having a nucleotide sequence which differs from those provided herein due to degeneracy of the genetic code are considered to be within the scope of the invention.

III. Additional Isolated Nucleic Acid Molecules of the Invention

In addition to isolated nucleic acids encoding alternative forms of T cell costimulatory molecules, the invention also discloses previously undescribed nucleotide sequences of the murine B7-1 gene and mRNA transcripts. As described in detail in Example 3, it has now been discovered that murine B7-1 mRNA transcripts contain additional 5' untranslated (UT) sequences which were not previously reported. A 5' UT region of approximately 250 base pairs has been reported for mB7-1 mRNA transcripts, determined by primer extension analysis (see Selvakumar et al. (1993) *Immunogenetics* 38:292-295). As described herein, an additional ~1500 nucleotides of 5' UT sequences have been discovered in mB7-1. These 5' UT sequences are contiguous with known exon 1 sequences, thereby extending the size of exon 1 by approximately 1500 base pairs. Thus the novel 5' UT sequence of the invention corresponds to the 5' region of mB7-1 exon 1 (i.e., exon 1 extends an additional ~1500 nucleotides at its 5' end than previously reported) rather than corresponding to a new exon upstream of exon 1. Computer analysis of the potential secondary structure of the 5' UT region reveals that the most stable structure is comprised of multiply folded palindromic sequences. This high degree of secondary structure may explain the results of Selvakumar et al. ((1993) *Immunogenetics* 38:292-295) in that the secondary structure could account for premature termination of the primer extension reaction. The potential for excessive secondary structure in the 5' UT region suggests that post-transcriptional mechanisms are involved in controlling mB7-1 expression. Thus, inclusion of the long 5' UT sequence in recombinant expression vectors encoding mB7-1 may provide post-transcriptional regulation that is similar to that of the endogenous gene. Accordingly, the 5' UT region of mB7-1 provided by the invention can be incorporated by standard recombinant DNA techniques at the 5' end of a cDNA encoding a mB7-1 protein. The nucleotide sequence of the 5' UT region of mB7-1 (i.e, the full nucleotide sequence of exon 1) is shown in SEQ ID NO: 6.

The discovery of additional 5' UT sequences in mB7-1 cDNA demonstrates that transcription of the mB7-1 gene initiates further upstream (i.e., 5') in genomic DNA than previously reported in Selvakumar et al. (*Immunogenetics* (1993) 38:292-295). Transcription of a gene is typically regulated by sequences in genomic DNA located immediately upstream of sequences corresponding to the 5' UT region of the transcribed mRNA. Nucleotides located within approximately 200 base pairs of the start site of transcription are generally considered to encompass the promoter of the gene and often include canonical CCAAT or TATA elements indicative of a typical eukaryotic promoter. For a gene having a promoter which contains a TATA box, transcription usually starts approximately 30 base pairs downstream of the TATA box. In addition to CCAAT and TATA-containing promoters, it is

molecule. For example, mRNA can be prepared from a sample of cells to be examined and the mRNA can be hybridized to an isolated nucleic acid encompassing a nucleotide sequence encoding all or a portion of an alternative cytoplasmic domain of a T cell costimulatory molecule (e.g., SEQ ID NO: 1) to detect the expression of the alternative cytoplasmic domain form of the costimulatory molecule in the cells. Furthermore, the isolated nucleic acids of the invention can be used to design oligonucleotide primers, e.g. PCR primers, which allow one to detect the expression of an alternatively spliced form of a T cell costimulatory molecule. Preferably, this oligonucleotide primer spans a novel exon junction created by alternative splicing and thus can only amplify cDNAs encoding this alternatively spliced form. For example, an oligonucleotide primer which spans exon 4 and exon 6 of murine B7-1 can be used to distinguish between the expression of a first cytoplasmic domain form of mB7-1 (i.e, encoded by exons 1-2-3-4-5) and expression of an alternative second cytoplasmic domain form of a costimulatory molecule (i.e., encoded by exons 1-2-3-4-6) (e.g., see Example 2).

The probes of the invention can be used to detect an alteration in the expression of an alternatively spliced form of a T cell costimulatory molecule, such as in a disease state. For example, detection of a defect in the expression of an alternatively spliced form of a T cell costimulatory molecule that is associated with an immunodeficiency disorder can be used to diagnose the disorder (i.e., the probes of the invention can be used for diagnostic purposes). Many congenital immunodeficiency diseases result from lack of expression of a cell-surface antigen important for interactions between T cells and antigen presenting cells. For example, the bare lymphocyte syndrome results from lack of expression of MHC class II antigens (see e.g., Rijkers, G.T. et al. (1987) *J. Clin. Immunol.* 7:98-106; Hume, C.R. et al. (1989) *Hum. Immunol.* 25:1-11)) and X-linked hyperglobulinemia results from defective expression of the ligand for CD40 (gp39) (see e.g. Korthauer, U et al. (1993) Nature 361:541; Aruffo, A. et al. (1993) *Cell* 72:291-300). An immunodeficiency disorder which results from lack of expression of an alternatively spliced form of a T cell costimulatory molecule can be diagnosed using a probe of the invention. For example, a disorder resulting from the lack of expression of the Cyt II form of B7-1 can be diagnosed in a patient based upon the inability of a probe which detects this form of B7-1 (e.g., an oligonucleotide spanning the junction of exon 4 and exon 6) to hybridize to mRNA in cells from the patient (e.g., by RT-PCR or by Northern blotting).

B. Recombinant Expression Vectors

An isolated nucleic acid of the invention can be incorporated into an expression vector (i.e., a recombinant expression vector) to direct expression of a novel structural form of a T cell costimulatory molecule encoded by the nucleic acid. The recombinant expression vectors are suitable for transformation of a host cell, and include a nucleic acid (or fragment thereof) of the invention and a regulatory sequence, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid. Operatively linked is

recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector (e.g., a nucleic acid of the invention) so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) *Nuc. Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques and are encompassed by the invention.

Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari. et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39).

Expression of alternatively spliced forms of T cell costimulatory molecules in mammalian cells is accomplished using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B., (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987), *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral material. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. The recombinant expression vector can be designed such that expression of the nucleic acid occurs preferentially in a particular cell type. In this situation, the expression vector's control functions are provided by regulatory sequences which allow for preferential expression of a nucleic acid contained in the vector in a particular cell type, thereby allowing for tissue or cell specific expression of an encoded protein.

The recombinant expression vectors of the invention can be a plasmid or virus, or viral portion which allows for expression of a nucleic acid introduced into the viral nucleic acid. For example, replication defective retroviruses, adenoviruses and adeno-associated viruses can be used. The recombinant expression vectors can be introduced into a host cell, e.g. *in vitro* or *in vivo*. A host cell line can be used to express a protein of the invention. Furthermore, introduction of a recombinant expression vector of the invention into a host cell can be used for therapeutic purposes when the host cell is defective in expressing the novel structural form of the T cell costimulatory molecule. For example, in a recombinant expression vector of the invention can be used for gene therapy purposes in a patient with an immunodeficiency disorder resulting from lack of expression of a novel structural form of a T cell costimulatory molecule.

C. Host Cells

The invention further provides a host cell transfected with a recombinant expression vector of the invention. The term "host cell" is intended to include prokaryotic and

cell is useful for studying signaling events and/or immunological responses which are mediated by the Cyt II domain rather than the Cyt I domain of B7-1. For example, one type of cell which can be used to create a host cell which exclusively expresses the Cyt II-form of murine B7-1 is a non-murine cell, since the non-murine cell does not express murine B7-1. Preferably, the non-murine cell also does not express other costimulatory molecules (e.g., COS cells can be used). Alternatively, a mouse cell which does not express the Cyt-I form of murine B7-1 can be used. For example, a recombinant expression vector of the invention can be introduced into NIH 3T3 fibroblast cells (which are B7-1 negative) or into cells derived from a mutant mouse in which the endogenous B7-1 gene has been disrupted and thus which does not natively express any form of B7-1 molecule (i.e., into cells derived from a "B7-1 knock-out" mouse, such as that described in Freeman, G.J. et al. (1993) *Science* 262:907-909).

In another embodiment, the host cell transfected with a recombinant expression vector encoding a novel structural form of a T cell costimulatory molecule is a tumor cell. Expression of the Cyt-I form of murine B7-1 on the surface of B7-1 negative murine tumor cells has been shown to induce T cell mediated specific immunity against the tumor cells accompanied by tumor rejection and prolonged protection to tumor challenge in mice (see Chen, L., et al. (1992) *Cell* 71, 1093-1102; Townsend, S.E. and Allison, J.P. (1993) *Science* 259, 368-370; Baskar, S., et al. (1993) *Proc. Natl. Acad. Sci.* 90, 5687-5690). Similarly, expression of novel structural forms of costimulatory molecules on the surface of a tumor cell may be useful for increasing the immunogenicity of the tumor cell. For example, tumor cells obtained from a patient can be transfected *ex vivo* with a recombinant expression vector of the invention, e.g., encoding an alternative cytoplasmic domain form of a costimulatory molecule, and the transfected tumor cells can then be returned to the patient. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*. Additionally, the tumor cell can also be transfected with recombinant expression vectors encoding other proteins to be expressed on the tumor cell surface to increase the immunogenicity of the tumor cell. For example, the Cyt-I form of B7-1, B7-2, MHC molecules (e.g., class I and/or class II) and/or adhesion molecules can be expressed on the tumor cells in conjunction with the Cyt-II form of B7-1.

D. Anti-Sense Nucleic Acid Molecules

The isolated nucleic acid molecules of the invention can also be used to design anti-sense nucleic acid molecules, or oligonucleotide fragments thereof, that can be used to modulate the expression of alternative forms of T cell costimulatory molecules. An anti-sense nucleic acid comprises a nucleotide sequence which is complementary to a coding strand of a nucleic acid, e.g. complementary to an mRNA sequence, constructed according to the rules of Watson and Crick base pairing, and can hydrogen bond to the coding strand of the nucleic acid. The hydrogen bonding of an antisense nucleic acid molecule to an mRNA

cytoplasmic domain (e.g. Cyt-II) can be made using the isolated nucleic acid shown in SEQ ID NO: 1 or SEQ ID NO: 3. Alternatively, a transgenic animal (e.g., a mouse) which expresses an mB7-2 protein containing an alternative signal peptide domain can be made using the isolated nucleic acid shown in SEQ ID NO: 12. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. These isolated nucleic acids can be linked to regulatory sequences which direct the expression of the encoded protein one or more particular cell types. Methods for generating transgenic animals, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009 and Hogan, B. et al., (1986) A Laboratory Manual, Cold Spring Harbor, New York, Cold Spring Harbor Laboratory. A transgenic founder animal can be used to breed additional animals carrying the transgene.

The isolated nucleic acids of the invention can further be used to create a non-human homologous recombinant animal. The term "homologous recombinant animal' as used herein is intended to describe an animal containing a gene which has been modified by homologous recombination. The homologous recombination event may completely disrupt the gene such that a functional gene product can no longer be produced (often referred to as a "knock-out" animal) or the homologous recombination event may modify the gene such that an altered, although still functional, gene product is produced. Preferably, the non-human animal is a mouse. For example, an isolated nucleic acid of the invention can be used to create a homologous recombinant mouse in which a recombination event has occurred in the B7-1 gene at an exon encoding a cytoplasmic domain such that this exon is altered (e.g., exon 5 or exon 6 is altered). Homologous recombinant mice can thus be created which express only the Cyt I or Cyt II domain form of B7-1. Accordingly, the invention provides a non-human knock-out animal which contains a gene encoding a B7-1 protein wherein an exon encoding a novel cytoplasmic domain is disrupted or altered.

To create an animal with homologously recombined nucleic acid, a vector is prepared which contains the DNA sequences which are to replace the endogenous DNA sequences, flanked by DNA sequences homologous to flanking endogenous DNA sequences (see for example Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected (see for example Li, E. et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see for example Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA.

embodiment, the isolated protein is a B7-1 or a B7-2 protein. E preferably comprises an amino acid sequence of a murine B7-1 cytoplasmic domain having an amino acid sequence shown in SEQ ID NO: 5 (i.e., the amino acid sequence of the cytoplasmic domain encoded by the novel exon 6 of the invention).

Another embodiment of the invention provides an isolated protein which binds CD28 or CTLA4 and is encoded by a T cell costimulatory molecule gene having at least one first exon encoding a first cytoplasmic domain and at least one second exon encoding a second cytoplasmic domain. The at least one first cytoplasmic domain comprises an amino acid sequence selected from the group consisting of amino acid sequence of SEQ ID NO:26 (mB7-1), SEQ ID NO:28 (hB7-1), SEQ ID NO:30 (mB7-2) and SEQ ID NO:32 (hB7-2). In this embodiment, the protein includes an amino acid sequence comprising at least one second cytoplasmic domain. Preferably, the protein does not include an amino acid sequence comprising a first cytoplasmic domain.

Preferred proteins which bind CD28 and/or CTLA4 are derived from B7-1 and B7-2. In a particularly preferred embodiment, the invention provides an isolated protein which binds CD28 or CTLA4 and has a novel cytoplasmic domain comprising an amino acid sequence shown in SEQ ID NO: 2.

A. Proteins with a Novel Signal Peptide Domain

In yet another aspect of the invention, T cell costimulatory molecules which include at least one novel signal peptide domain are provided. In one embodiment, the isolated protein binds to CD28 or CTLA4 and has an amino acid sequence derived from amino acid sequences encoded by at least one T cell costimulatory molecule gene. In this embodiment, the protein comprises a contiguous amino acid sequence represented by a formula A-B-C-D-E, wherein

A comprises an amino acid sequence of a signal peptide domain encoded by at least one exon of a T cell costimulatory molecule gene,

B comprises an amino acid sequence of an immunoglobulin variable region-like domain encoded by at least one exon of a T cell costimulatory molecule gene,

C comprises an amino acid sequence of an immunoglobulin constant region-like domain encoded by at least one exon of aT cell costimulatory molecule gene,

D, which may or may not be present, comprises an amino acid sequence of a transmembrane domain encoded by at least one exon of a T cell costimulatory molecule gene, and

E, which may or may not be present, comprises an amino acid sequence of a cytoplasmic domain encoded by at least one exon of a T cell costimulatory molecule gene,

In the formula, A, B, C and D are contiguous amino acid residues linked by amide bonds from an N-terminus to a C-terminus. In a preferred embodiment, an isolated murine B7-1 protein having an IgV-like domain deleted comprises an amino acid sequence shown in SEQ ID NO: 9 (utilizing Cyt I of mB7-1). Alternatively, an isolated murine B7-1 protein having an IgV-like domain deleted comprises an amino acid sequence shown in SEQ ID NO: 11 (utilizing Cyt II of mB7-1).

In another embodiment, the structural form of the T cell costimulatory molecule has at least one IgC-like domain deleted. Accordingly, in one embodiment, the isolated protein has an amino acid sequence derived from amino acid sequences encoded by at least one T cell costimulatory molecule gene and comprises a contiguous amino acid sequence represented by a formula A-B-C-D, wherein

A, which may or may not be present, comprises an amino acid sequence of a signal peptide domain encoded by at least one exon of a T cell costimulatory molecule gene,

B comprises an amino acid sequence of an immunoglobulin variable region-like domain encoded by at least one exon of a T cell costimulatory molecule gene, and

C comprises an amino acid sequence of a transmembrane domain encoded by at least one exon of a T cell costimulatory molecule gene, and

D comprises an amino acid sequence of a cytoplasmic domain encoded by at least one exon of a T cell costimulatory molecule gene.

In the formula, A, B, C and D are contiguous amino acid residues linked by amide bonds from an N-terminus to a C-terminus. In a preferred embodiment, an isolated murine B7-1 protein having an IgC-like domain deleted comprises an amino acid sequence shown in SEQ ID NO: 63 (utilizing Cyt I of mB7-1). Alternatively, an isolated murine B7-1 protein having an IgC-like domain deleted comprises an amino acid sequence shown in SEQ ID NO: 65 (utilizing Cyt II of mB7-1).

The proteins of the invention can be isolated by expression of the molecules (e.g., proteins or peptide fragments thereof) in a suitable host cell using techniques known in the art. Suitable host cells include prokaryotic or eukaryotic organisms or cell lines, for example, yeast, *E. coli* and insect cells. The recombinant expression vectors of the invention, described above, can be used to express a costimulatory molecule in a host cell in order to isolate the protein. The invention provides a method of preparing an isolated protein of the invention comprising introducing into a host cell a recombinant expression vector encoding the protein, allowing the protein to be expressed in the host cell and isolating the protein. Proteins can be isolated from a host cell expressing the protein according to standard procedures of the art, including ammonium sulfate precipitation, fractionation column

T cell costimulatory molecule can be used to inhibit a costimulatory signal in T cells by contacting the T cells with the soluble molecule.

B. Antibodies

A novel structural form of a T cell costimulatory molecule of the invention can be used to produce antibodies directed against the costimulatory molecule. Conventional methods can be used to prepare the antibodies. For example, to produce polyclonal antibodies, a mammal, (e.g., a mouse, hamster, or rabbit) can be immunized with a costimulatory molecule, or an immunogenic portion thereof, which elicits an antibody response in the mammal. Techniques for conferring immunogenicity on a protein include conjugation to carriers or other techniques well known in the art. For example, the protein can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassay can be used with the immunogen as antigen to assess the levels of antibodies. Following immunization, antisera can be obtained and, if desired, polyclonal antibodies isolated from the sera.

In addition to polyclonal antisera, the novel costimulatory molecules of the invention can be used to raise monoclonal antibodies. To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused with myeloma cells by standard somatic cell fusion procedures thus immortalizing these cells and yielding hybridoma cells. Such techniques are well known in the art. For example, the hybridoma technique originally developed by Kohler and Milstein (*Nature* 256, 495-497 (1975)) as well as other techniques such as the human B-cell hybridoma technique (Kozbor et al., *Immunol. Today* 4, 72 (1983)), the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al. *Monoclonal Antibodies in Cancer Therapy* (1985) Allen R. Bliss, Inc., pages 77-96), and screening of combinatorial antibody libraries (Huse et al., *Science* 246, 1275 (1989)). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with the protein or portion thereof and monoclonal antibodies isolated.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with an alternative cytoplasmic domain of a costimulatory molecule. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments.

Chimeric and humanized antibodies are also within the scope of the invention. It is expected that chimeric and humanized antibodies would be less immunogenic in a human subject than the corresponding non-chimeric antibody. A variety of approaches for making chimeric antibodies, comprising for example a non-human variable region and a human

costimulatory molecule can be stimulated through the costimulatory molecule, e.g., by crosslinking the costimulatory molecules on the cell surface with an antibody, and intracellular signals and/or other cellular changes (e.g., changes in surface expression of proteins etc.) induced thereupon can be identified.

Additionally, an isolated T cell costimulatory molecule of the invention comprising a novel cytoplasmic domain can be used in methods of identifying other molecules (e.g., proteins) which interact with (i.e., bind to) the costimulatory molecule using standard *in vitro* assays (e.g., incubating the isolated costimulatory molecule with a cellular extract and determining by immunoprecipitation if any molecules within the cellular extract bind to the costimulatory molecule). It is of particular interest to identify molecules which can interact with the novel cytoplasmic domain since such molecules may also be involved in intracellular signaling. For example, it is known that the cytoplasmic domains of many cell-surface receptors can interact intracellularly with other members of the signal transduction machinery, e.g., tyrosine kinases.

The invention further provides a method for screening agents to identify an agent which upregulates or downregulates expression of a novel structural domain form of a T cell costimulatory molecule. The method involves contacting a cell which expresses or can be induced to express a T cell costimulatory molecule with an agent to be tested and determining expression of a novel structural domain form of the T cell costimulatory molecule by the cell. The term "upregulates" encompasses inducing the expression of a novel form of a T cell costimulatory molecule by a cell which does not constitutively express such a molecule or increasing the level of expression of a novel form of a T cell costimulatory molecule by a cell which already expresses such a molecule. The term "downregulates" encompasses decreasing or eliminating expression of an a novel form of a T cell costimulatory molecule by a cell which already expresses such a molecule. The term "agent" is intended to include molecules which trigger an upregulatory or downregulatory response in a cell. For example, an agent can be a small organic molecule, a biological response modifier (e.g., a cytokine) or a molecule which can crosslink surface structures on the cell (e.g., an antibody). For example, expression of the a novel cytoplasmic domain form of the T cell costimulatory molecule by the cell can be determined by detecting an mRNA transcript encoding the novel cytoplasmic domain form of the T cell costimulatory molecule in the cell. For example, mRNA from the cell can be reverse transcribed and used as a template in PCR reactions utilizing PCR primers which can distinguish between a Cyt I cytoplasmic domain form and a novel Cyt II cytoplasmic domain form of the T cell costimulatory molecule (see e.g., Example 2). Alternatively, a novel cytoplasmic domain-containing T cell costimulatory molecule can be detected in the cell using an antibody directed against the novel cytoplasmic domain (e.g., by immunoprecipitation or immunohistochemistry). A preferred T cell costimulatory molecule for use in the method is B7-1. Cell types which are known to express the Cyt-I form of B7-1, or which can be induced to express the Cyt-I form of B7-1, include B

A fusion protein of the invention, comprising a first peptide fused to a second peptide comprising a novel cytoplasmic domain of the invention, can be used to transfer the signal transduction function of the novel cytoplasmic domain to another protein. For example, a novel cytoplasmic domain of B7-1 (e.g., Cyt-II) can be fused to the extracellular and transmembrane domains of another protein (e.g., an immunoglobulin protein, a T cell receptor protein, a growth factor receptor protein etc.) and the fusion protein can be expressed in a host cell by standard techniques. The extracellular domain of the fusion protein can be crosslinked (e.g., by binding of a ligand or antibody to the extracellular domain) to generate an intracellular signal(s) mediated by the novel cytoplasmic domain.

Additionally, a fusion protein of the invention can be used in methods of identifying and isolating other molecules (e.g., proteins) which can interact intracellularly (i.e., within the cell cytoplasm) with a novel cytoplasmic domain of the invention. One approach to identifying molecules which interact intracellularly with the cytoplasmic domain of a cell-surface receptor is to metabolically label cells which express the receptor, immunoprecipitate the receptor, usually with an antibody against the extracellular domain of the receptor, and identify molecules which are co-immunoprecipitated along with the receptor. In the case of mB7-1, however, the cells which have been found to express the naturally-occurring Cyt-II form of B7-1 have also been found to express the naturally-occurring Cyt-I form of B7-1 (e.g., thymocytes, see Example 2). Thus, immunoprecipitation with an antibody against the extracellular domain of mB7-1 would immunoprecipitate both forms of the protein since the extracellular domain is common to both the Cyt-I and Cyt-II containing forms. Thus, molecules which interact with either Cyt-I or Cyt-II would be co-immunoprecipitate. A fusion protein comprising a non-B7-1 extracellular domain (to which an antibody can bind), a transmembrane domain (derived either from the non-B7-1 molecule or from B7-1) and a B7-1 alternative cytoplasmic domain (e.g., Cyt-II) can be constructed and expressed in a host cell which naturally expresses the Cyt-II form of B7-1. The antibody directed against the "heterologous" extracellular domain of the fusion protein can then be used to immunoprecipitate the fusion protein and to co-immunoprecipitate any other proteins which interact intracellularly with the novel cytoplasmic domain.

B. Antibodies

An antibody which binds to a novel structural domain of the invention can be prepared by using the domain, or a portion thereof, as an immunogen. Polyclonal antibodies or monoclonal antibodies can be prepared by standard techniques described above. In a preferred approach, peptides comprising amino acid sequences of the domain are used as immunogens, e.g. overlapping peptides encompassing the amino acid sequence of the domain. For example, polyclonal antisera against a novel cytoplasmic domain (e.g., Cyt II of mB7-1) can be made by preparing overlapping peptides encompassing the amino acid

Genomic cloning

A mouse 129 lambda genomic library was kindly provided by Drs. Hong Wu and Rudolf Jaenisch of the Whitehead Institute for Biomedical Research, Cambridge, MA. Genomic DNA was prepared from the J1 embryonic stem cell line (derived from the 129/sv mouse strain), partially digested with MboI, sized (17-21 kb), and ligated into the BamHI site of lambda-DASH II arms (Stratagene, La Jolla CA). The library was probed with the coding region of mB7-1 cDNA to yield four clones (λ4, λ9, λ15, and λ16). These lambda clones were subcloned into Bluescript-pKS II (Stratagene, La Jolla CA) for subsequent restriction mapping.

Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

Total cellular RNA was prepared from SWR/J mouse spleen and thymus using RNA-Stat-60 (Tel-Test "B", Inc, Friendswood, Texas). Random hexamer primed reverse transcription (RT) was performed with Superscript-RT (Gibco BRL, Gaithersburg MD) using 1-10 μg total RNA in a 20 μl reaction. All PCR reactions were performed in 25 μl volumes using a manual "hot start", wherein 10X deoxynucleotide triphosphates (dNTPs) were added to the samples at 80 °C. Final reaction conditions were: 60 mM Tris-HCl, pH 8.5, 15 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 200 μM dNTPs, and 2 μg/ml each of the specific primers. Cycling conditions for all amplifications were 94° C, 4 minutes prior to 35 cycles of 94° C for 45 seconds, 58° C for 45 seconds, and 72° C for 3 minutes, followed by a final extension at 72° C for 7 minutes. The template for primary PCR was 2 μl of the RT reaction product and the template for secondary nested PCR was 1 μl of the primary PCR reaction product.

Oligonucleotides

All oligonucleotides were synthesized on an Applied Biosystems 381A DNA Synthesizer. The oligonucleotides used in this study are listed in Table I and their uses for primary or secondary PCR, as well as sense, also are indicated.

Rapid Amplification of cDNA Ends (RACE) Procedure

Polyadenylated RNA purified by two cycles of oligo-dT selection was obtained from CH1 B lymphoma cells, which express high levels of mB7-1. Primers designed to the most 5' end of the cDNA were employed with the 5' RACE Kit (Gibco BRL, Gaithersburg, MD) according to the manufacturer's instructions. In brief, RNA was reverse transcribed with a gene-specific oligonucleotide, the cDNA purified, and a poly-dCTP tail was added with terminal deoxynucleotide transferase. PCR was performed using a nested primer and an oligonucleotide complimentary to the poly-dCTP tail. PCR bands were cloned, sequenced, and correlated with the genomic sequences.

EXAMPLE 2: Identification of mB7-1 exon 6: An alternately spliced exon encoding a novel second cytoplasmic domain

Analysis of mB7-1 cDNAs isolated from an A20 B cell cDNA library showed that one cDNA contained additional sequence not previously described for the mB7-1 cDNA. This sequence was mapped to the mB7-1 locus approximately 7-kb downstream of exon 5. A canonical splice site was present immediately upstream of this sequence and a poly-adenylation site was present downstream. Taken together, these data suggested that this novel sequence represents an additional exon, encoding 46 amino acids, which may be alternatively spliced in place of exon 5. This alternative cytoplasmic domain is notable for two casein kinase II phosphorylation sites (amino acid positions 11-15 (SAKDF) and amino acid positions 28-32 (SLGEA) of SEQ ID NO: 5) (for a description of casein kinase II phosphorylation sites see Pinna (1990) *Biochimica et Biophysica Acta* 1054:267-284) and one protein kinase C phosphorylation site (amino acid positions 11-14 (SAKD) of SEQ ID NO: 5)(for a description of protein kinase C phosphorylation sites see Woodgett et al. (1986) *Biochemistry* 161:177-184; and Kishimoto et al. (1985) *J. Biol. Chem.* 260:12492-12499).

In order to assess whether exon 6 also could be used in an alternative fashion, an antisense primer (B7.48) was designed to the predicted exon 4/6 splice junction such that only the alternatively spliced product would give rise to an amplified product. This primer overhangs the putative exon 4/6 junction by 3 bp at its 3' end. The 3 bp overhang is insufficient to permit direct priming in exon 4 outside the context of an exon 4/6 splice (Figure 1, lane 9, negative control is a cDNA clone containing only mB7-1 CytI). The expected amplified product for the alternately spliced transcript (Figure 1, transcript C) would be 399 bp. Interestingly, this transcript was observed only in thymic, but not splenic RNA.

[In Figure1, lanes 1, 2 and 3 represent nested PCR products from murine splenic RNA using PCR primers B7.27-B7.36, B7.27-B7.38, and B7.27-B7.48, respectively. Lanes 4, 5 and 6 represent nested PCR products from murine thymic RNA using PCR primers B7.27-B7.36, B7.27-B7.38 and B7.27-B7.48, respectively. Lane 7 represents a negative control (no input RNA). Lane 8 represents a positive control (mB7-1 cDNA clone). Lane 9 represents a negative control for B7.27-B7.48 amplification comprised of the mB7-1 cDNA containing cytoplasmic domain I, which does not have the correct exon 4-6 splice junction. Lane M is a 100 bp ladder with the lower bright band equal to 600 bp. Letters A, B and C refer to the transcripts detected and are further illustrated in Figure 1. Note that exon 6 splicing as an alternative cytoplasmic domain is present only in the thymus, but not in the spleen].

To further investigate the use of exon 6 in mB7-1 mRNA transcripts, nested RT-PCR spanning exons 3 through 6 was performed using spleen RNA (Figure 1, PCR product A). A PCR product longer than predicted from the use of exon 6 as an alternatively spliced exon also was observed. Subsequent sequence analysis indicated that in this transcript, exons 5 and 6 were spliced in tandem, rather than in an alternative fashion (Figure 1, transcript A),

junctions have been reported previously (Selvakumar et al. (1993) *Immunogenetics* 38:292-295). The coding region of the exon 1 signal peptide domain is 115 bp and is flanked at the 3' end with a canonical splice site. Exons 2 (318 bp), 3 (282 bp), and 4 (114 bp), are separated by 6.0 and 3.8 kb, respectively, and all 3 exons are flanked on both their 5' and 3' ends with canonical splice sites. Exon 5 is located 4 kb downstream of exon 4, and contains a termination codon after the first 97 bp. An additional functional canonical splice site was observed 43 bp downstream of the termination codon in exon 5, since this site was used to generate the transcript outlined in Figure 1 (transcript A). Exon 6 is located 7.2 kb downstream of exon 5 and encodes an open reading frame with a termination codon after 140 bp. Both exons 5 and 6 are followed by polyadenylation sequences, ATTAAA and AATAAA respectively.

EXAMPLE 5: Identification of Additional Novel Cytoplasmic Domains by Exon Trapping

In this example, an exon trapping approach is used to identify a novel exon encoding an alternative cytoplasmic domain for human B7-1. The basic strategy of exon trapping is to create an expression vector encoding a recombinant protein, wherein the encoded protein cannot be functionally expressed unless an appropriate exon, with flanking intron sequences that allow proper mRNA splicing, is cloned into the expression vector. A recombinant expression vector is created comprising transcriptional regulatory sequences (e.g., a strong promoter) linked to nucleic acid encoding the human B7-1 signal peptide exon, IgV-like and IgC-like exons followed by a transmembrane exon with flanking 3' intron donor splice sequences. These splice sequences are immediately followed by translational stop codons in all three frames. A polyadenylation recognition site is not included in the recombinant expression vector. Following the stop codons are restriction enzyme sites which allow genomic DNA fragments to be cloned into the expression vector to create a library of recombinant expression vectors.

As a negative control, the parental recombinant expression vector is transfected into a host cell line which is hB7-1$^-$ (e.g, COS cells) and the absence of surface expression of hB7-1 is demonstrated, confirming that the parental expression vector alone is unable to direct stable surface expression of hB7-1 in the absence of a cytoplasmic domain encoding exon. As a positive control, the known hB7-1 cytoplasmic domain with a flanking 5' intron acceptor splice sequence is cloned into a restriction enzyme site downstream of the transmembrane exon such that the transmembrane domain exon can be spliced to the cytoplasmic domain exon. This positive control vector is transfected into a host cell (e.g., COS cells) and the surface expression of hB7-1 on the cells is demonstrated, confirming that the cloning into the vector of a cytoplasmic domain encoding exon with the proper splice sequences produces an hB7-1 molecule that can be stably expressed on the cell surface.

complementary to the poly-dCTP tail to amplify 5' cDNA fragments of mB7-2 transcripts. The gene-specific oligonucleotide primers used for PCR were as follows:

CAGCTCACTCAGGCTTATGT	reverse transcription, - sense	(SEQ ID NO: 55)
AAACAGCATCTGAGATCAGCA	primary PCR, - sense	(SEQ ID NO: 56)
CTGAGATCAGCAAGACTGTC	secondary PCR, - sense	(SEQ ID NO: 57)

The amplified fragments were subcloned into a plasmid vector and sequenced. Of approximately 100 individual clones examined, ~75 % of the clones had a 5' nucleotide sequence corresponding to that reported for the 5' end of an mB7-2 cDNA (see Freeman, G.J. et al. (1993) *J. Exp. Med.* 178:2185-2192). Approximately 25 % of the clones had a 5' nucleotide sequence shown in SEQ ID NO:14, which encodes a novel signal peptide domain having an amino acid sequence shown in SEQ ID NO:15.

EXAMPLE 7: Identification of Alternatively Spliced Forms of B7-1 Having a Structural Domain Deleted

Reverse-transcriptase polymerase chain reaction was used to amplify mB7-1 cDNA fragments derived from murine spleen cell RNA. Oligonucleotide primers used for PCR were as follows:

CTGAAGCTATGGCTTGCAATT	primary PCR, + sense	(SEQ ID NO: 58)
ACAAGTGTCTTCAGATGTTGAT	secondary PCR, + sense	(SEQ ID NO: 59)
CTGGATTCTGACTCACCTTCA	primary PCR, - sense	(SEQ ID NO: 60)
CCAGGTGAAGTCCTCTGACA	secondary PCR, - sense	(SEQ ID NO: 61)

A cDNA fragment was detected which comprises a nucleotide sequence (SEQ ID NO:8) encoding a murine B7-1 molecule in which the signal peptide domain was spliced directly to the IgC-like domain (i.e., the IgV-like domain was deleted). The amino acid sequence of mB7-1 encoded by this cDNA is shown in SEQ ID NO:9.

Another cDNA fragment was detected with comprises a nucleotide sequence (SEQ ID NO: 62) encoding a murine B7-1 molecule in which the IgV-like domain was spliced directly to the transmembrane domain (i.e., the IgC-like domain was deleted). The amino acid sequence encoded by this cDNA is shown in SEQ ID NO: 63). This protein is referred to herein as an IgV-like isoform of mB7-1. To examine the functional activity of the IgV-like

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: BRIGHAM AND WOMEN'S HOSPITAL
(B) STREET: 75 FRANCIS STREET
(C) CITY: BOSTON
(D) STATE: MASSACHUSETTS
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02115
(A) NAME: DANA-FARBER CANCER INSTITUTE
(B) STREET: 44 BINNEY STREET
(C) CITY: BOSTON
(D) STATE: MASSACHUSETTS
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02115

(ii) TITLE OF INVENTION: Novel Forms of T Cell Costimulatory Molecules and Uses Therefor

(iii) NUMBER OF SEQUENCES: 65

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LAHIVE & COCKFIELD
(B) STREET: 60 State Street, suite 510
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02109-1875

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/205,697
(B) FILING DATE: 02-Mar-1994

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Mandragouras, Amy E.
(B) REGISTRATION NUMBER: 36,207
(C) REFERENCE/DOCKET NUMBER: BWI-120CPPC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)227-5941

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1888 base pairs

```
GCT GAC TTC TCT ACC CCC AAC ATA ACT GAG TCT GGA AAC CCA TCT GCA      722
Ala Asp Phe Ser Thr Pro Asn Ile Thr Glu Ser Gly Asn Pro Ser Ala
        145                 150                 155

GAC ACT AAA AGG ATT ACC TGC TTT GCT TCC GGG GGT TTC CCA AAG CCT      770
Asp Thr Lys Arg Ile Thr Cys Phe Ala Ser Gly Gly Phe Pro Lys Pro
    160                 165                 170

CGC TTC TCT TGG TTG GAA AAT GGA AGA GAA TTA CCT GGC ATC AAT ACG      818
Arg Phe Ser Trp Leu Glu Asn Gly Arg Glu Leu Pro Gly Ile Asn Thr
175                 180                 185                 190

ACA ATT TCC CAG GAT CCT GAA TCT GAA TTG TAC ACC ATT AGT AGC CAA      866
Thr Ile Ser Gln Asp Pro Glu Ser Glu Leu Tyr Thr Ile Ser Ser Gln
                195                 200                 205

CTA GAT TTC AAT ACG ACT CGC AAC CAC ACC ATT AAG TGT CTC ATT AAA      914
Leu Asp Phe Asn Thr Thr Arg Asn His Thr Ile Lys Cys Leu Ile Lys
            210                 215                 220

TAT GGA GAT GCT CAC GTG TCA GAG GAC TTC ACC TGG GAA AAA CCC CCA      962
Tyr Gly Asp Ala His Val Ser Glu Asp Phe Thr Trp Glu Lys Pro Pro
        225                 230                 235

GAA GAC CCT CCT GAT AGC AAG AAC ACA CTT GTG CTC TTT GGG GCA GGA     1010
Glu Asp Pro Pro Asp Ser Lys Asn Thr Leu Val Leu Phe Gly Ala Gly
    240                 245                 250

TTC GGC GCA GTA ATA ACA GTC GTC GTC ATC GTT GTC ATC ATC AAA TGC     1058
Phe Gly Ala Val Ile Thr Val Val Val Ile Val Val Ile Ile Lys Cys
255                 260                 265                 270

TTC TGT AAG CAC GGT CTC ATC TAC CAT TTG CAA CTG ACC TCT TCT GCA     1106
Phe Cys Lys His Gly Leu Ile Tyr His Leu Gln Leu Thr Ser Ser Ala
                275                 280                 285

AAG GAC TTC AGA AAC CTA GCA CTA CCC TGG CTC TGC AAA CAC GGT TCT     1154
Lys Asp Phe Arg Asn Leu Ala Leu Pro Trp Leu Cys Lys His Gly Ser
            290                 295                 300

CTA GGT GAA GCC TCT GCA GTG ATT TGC AGA AGT ACT CAG ACG AAT GAA     1202
Leu Gly Glu Ala Ser Ala Val Ile Cys Arg Ser Thr Gln Thr Asn Glu
        305                 310                 315

CCA CAG TAGTTCTGCT GTTTCTGAGG ACGTAGTTTA GAGACTGAAT TCTTTGGAAA     1258
Pro Gln
    320

GGACATAGGG ACAGTTTGCA CATTTGCTTG CACATCACAC ACACACACAC ACACACACAC   1318

ACACACACAC ACACACACAC ACACACACAC ACACACACAC TCTCTCTCTC TCTCTCTCTC   1378

GATACCTTAG GATAGGGTTC TACCCTGTTG CTCAGTGACA AAGAATCACT CTGTGGCGGA   1438

GGCAGGCTTC AAGCTTGCAG CAATCCTCCT GCACCAGTTT CCTGAGTGCC AGACTTCCAG   1498

GTGTAAGCTA TGGCACTTAG CAGAACACTA GCTGAATCAA TGAAGACACT GAGGTTCCAA   1558

GAGGGAACCT GAATTATGAA GGTGAGTCAG AATCCAGATT TCCTGGCTCT ACCACTCTTA   1618
```

Ser Gln Asp Pro Glu Ser Glu Leu Tyr Thr Ile Ser Ser Gln Leu Asp
195 200 205

Phe Asn Thr Thr Arg Asn His Thr Ile Lys Cys Leu Ile Lys Tyr Gly
210 215 220

Asp Ala His Val Ser Glu Asp Phe Thr Trp Glu Lys Pro Pro Glu Asp
225 230 235 240

Pro Pro Asp Ser Lys Asn Thr Leu Val Leu Phe Gly Ala Gly Phe Gly
245 250 255

Ala Val Ile Thr Val Val Val Ile Val Val Ile Ile Lys Cys Phe Cys
260 265 270

Lys His Gly Leu Ile Tyr His Leu Gln Leu Thr Ser Ser Ala Lys Asp
275 280 285

Phe Arg Asn Leu Ala Leu Pro Trp Leu Cys Lys His Gly Ser Leu Gly
290 295 300

Glu Ala Ser Ala Val Ile Cys Arg Ser Thr Gln Thr Asn Glu Pro Gln
305 310 315 320

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 249..1166

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAGTTTTATA CCTCAATAGA CTCTTACTAG TTTCTCTTTT TCAGGTTGTG AAACTCAACC 60

TTCAAAGACA CTCTGTTCCA TTTCTGTGGA CTAATAGGAT CATCTTTAGC ATCTGCCGGG 120

TGGATGCCAT CCAGGCTTCT TTTTCTACAT CTCTGTTTCT CGATTTTTGT GAGCCTAGGA 180

GGTGCCTAAG CTCCATTGGC TCTAGATTCC TGGCTTTCCC CATCATGTTC TCCAAAGCAT 240

CTGAAGCT ATG GCT TGC AAT TGT CAG TTG ATG CAG GAT ACA CCA CTC CTC 290
Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu
1 5 10

AAG TTT CCA TGT CCA AGG CTC AAT CTT CTC TTT GTG CTG CTG AAT CGT 338
Lys Phe Pro Cys Pro Arg Leu Asn Leu Leu Phe Val Leu Leu Asn Arg
15 20 25 30

TTC TGT AAG CAC AGA AGC TGT TTC AGA AGA AAT GAG GCA AGC AGA GAA 1106
Phe Cys Lys His Arg Ser Cys Phe Arg Arg Asn Glu Ala Ser Arg Glu
275 280 285

ACA AAC AAC AGC CTT ACC TTC GGG CCT GAA GAA GCA TTA GCT GAA CAG 1154
Thr Asn Asn Ser Leu Thr Phe Gly Pro Glu Glu Ala Leu Ala Glu Gln
290 295 300

ACC GTC TTC CTT TAGTTCTTCT CTGTCCATGT GGGATACATG GTATTATGTG 1206
Thr Val Phe Leu
305

GCTCATGAGG TACAATCTTT CTTTCAGCAC CGTGCTAGCT GATCTTTCGG ACAACTTGAC 1266

ACAAGATAGA GTTAACTGGG AAGAGAAAGC CTTGAATGAG GATTTCTTTC CATCAGGAAG 1326

CTACGGGCAA GTTTGCTGGG CCTTTGATTG CTTGATGACT GAAGTGGAAA GGCTGAGCCC 1386

ACTGTGGGTG GTGCTAGCCC TGGGCAGGGG CAGGTGACCC TGGGTGGTAT AAGAAAAAGA 1446

GCTGTCACTA AAAGGAGAGG TGCCTAGTCT TACTGCAACT TGATATGTCA TGTTTGGTTG 1506

GTGTCTGTGG GAGGCCTGCC CTTTTCTGAA GAGAAGTGGT GGGAGAGTGG ATGGGGTGGG 1566

GGCAGAGGAA AAGTGGGGGA GAGGGCCTGG GAGGAGAGGA GGGAGGGGGA CGGGGTGGGG 1626

GTGGGGAAAA CTATGGTTGG GATGTAAAAA CGGATAATAA TATAAATATT AAATAAAAAG 1686

AGAGTATTGA GCGGTCTCAT CTACCATTTG CAACTGACCT CTTCTGCAAA GGACTTCAGA 1746

AACCTAGCAC TACCCTGGCT CTGCAAACAC GGTTCTCTAG GTGAAGCCTC TGCAGTGATT 1806

TGCAGAAGTA CTCAGACGAA TGAACCACAG TAGTTCTGCT GTTTCTGAGG ACGTAGTTTA 1866

GAGACTGAAT TCTTTGGAAA GGACATAGGG ACAGTTTGCA CATTTGCTTG CACATCACAC 1926

ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC 1986

TCTCTCTCTC TCTCTCTCTC GATACCTTAG GATAGGGTTC TACCCTGTTG CTCAGTGACA 2046

AAGAATCACT CTGTGGCGGA GGCAGGCTTC AAGCTTGCAG CAATCCTCCT GCACCAGTTT 2106

CCTGAGTGCC AGACTTCCAG GTGTAAGCTA TGGCACTTAG CAGAACACTA GCTGAATCAA 2166

TGAAGACACT GAGGTTCCAA GAGGGAACCT GAATTATGAA GGTGAGTCAG AATCCAGATT 2226

TCCTGGCTCT ACCACTCTTA ACCTGTATCT GTTAGACCCC AAGCTCTGAG CTCATAGACA 2286

AGCTAATTTA AAATGCTTTT TAATAAGCAG AAGGCTCAGT TAGTACGGGG TTCAGGATAC 2346

TGCTTACTGG CAATATTTGA CTAGCCTCTA TTTTGTTTGT TTTTTAAAGG CCTACTGACT 2406

GTAGTGTAAT TTGTAGGAAA CATGTTGCTA TGTATACCCA TTTGAGGGTA ATAAAAATGT 2466

TGGTAATTTT CAGCCAGCAC TTTCCAGGTA TTTCCCTTTT TATCCTTCAT 2516

(2) INFORMATION FOR SEQ ID NO:4:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gly Leu Ile Tyr His Leu Gln Leu Thr Ser Ser Ala Lys Asp Phe Arg
1 5 10 15

Asn Leu Ala Leu Pro Trp Leu Cys Lys His Gly Ser Leu Gly Glu Ala
20 25 30

Ser Ala Val Ile Cys Arg Ser Thr Gln Thr Asn Glu Pro Gln
35 40 45

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1753 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTTTTAGTAA CCAGAGGCCG CAAGAAGAGA TCACTTGTAT ATACACGGGC CCCATCTTTT 60

GCTTTTTAAG ACAAAAGAAA AAGAATCTTC TTCAACAAGT AAGTAAATGC ATTTACTATT 120

TATCATGCTA TGGGACACCT TAGTAGAACA CGCTATCTCC AGCCTTATCA TATGCATATT 180

TTGTTGTTGT TGTTGTTGTT GTTGTTAAAG ACAGGGTCTC ATATATGCCA GGCTGGTCCC 240

AAACTTTCAG TGTAACCCAA GATAATCTGG AACTCCCGAC TCCTCTGCTC CCACCTCTCC 300

AGTGCAGGAC ACTGTTTATA CCGTGCTGGG GAATTGAACT CAGAGCACCC TGCATGTCAG 360

CTAAGCATTC TACCGACCAA GTCCCATGCC CAGTCCCTAA CTCCCCAACT TCACTGCTTT 420

TTAAACATAC ATACAATCAT AACTTGCCCT CAGAGCAGTC TCCTGGGGTC TCTTATTCTC 480

AAGGCTGCGG CATTCCAACA CTGTTAGAAA AACACCATCA GGATTCTTTT GTGTTTCCTA 540

GATGCAAACA TTTTTGTAGG GCGAAGTTGA GGTTTTTCTA ATCAAGAAAA TGCCGGTAAC 600

AAGTCTCTTC AAGCTAACTG GTTGGCTAAG GGGTATCTCT CCAAAAGAAG AGATCCACAT 660

GTCAGGCCAG TTGTAGGCAT GATGTCAGGT CTCCCTCCCT TTCTTTCTTT CTTTCTTTTT 720

TTCTTTCTTT CTTTTTTTCT TTCTTTCTTA CTTTCTTACT TTCTTTCTTT TCTGTTTTTT 780

GGTTTTTCGA GACAGGGTTT CTTTGTATAG CCCTGGCTGT CCTGGAACTC GCTCTGTAGA 840

CCAGGCTGGC CTCGAACTCA GAAATCTGCC TCTGCCTTTA CCTCCTGAGT GCTGGGAATT 900

AAAGGTGTGC ACCACCATGC CCGGCTGGGA TGTCATTCGT TTTCATTTCT CAATTTTGAT 960

(B) LOCATION: 249..848

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
GAGTTTTATA CCTCAATAGA CTCTTACTAG TTTCTCTTTT TCAGGTTGTG AAACTCAACC      60

TTCAAAGACA CTCTGTTCCA TTTCTGTGGA CTAATAGGAT CATCTTTAGC ATCTGCCGGG     120

TGGATGCCAT CCAGGCTTCT TTTTCTACAT CTCTGTTTCT CGATTTTTGT GAGCCTAGGA     180

GGTGCCTAAG CTCCATTGGC TCTAGATTCC TGGCTTTCCC CATCATGTTC TCCAAAGCAT     240

CTGAAGCT ATG GCT TGC AAT TGT CAG TTG ATG CAG GAT ACA CCA CTC CTC       290
         Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu
           1               5                  10

AAG TTT CCA TGT CCA AGG CTC AAT CTT CTC TTT GTG CTG CTG ATT CGT       338
Lys Phe Pro Cys Pro Arg Leu Asn Leu Leu Phe Val Leu Leu Ile Arg
 15                  20                  25                  30

CTT TCA CAA GTG TCT TCA GCT GAC TTC TCT ACC CCC AAC ATA ACT GAG       386
Leu Ser Gln Val Ser Ser Ala Asp Phe Ser Thr Pro Asn Ile Thr Glu
                 35                  40                  45

TCT GGA AAC CCA TCT GCA GAC ACT AAA AGG ATT ACC TGC TTT GCT TCC       434
Ser Gly Asn Pro Ser Ala Asp Thr Lys Arg Ile Thr Cys Phe Ala Ser
             50                  55                  60

GGG GGT TTC CCA AAG CCT CGC TTC TCT TGG TGG GAA AAT GGA AGA GAA       482
Gly Gly Phe Pro Lys Pro Arg Phe Ser Trp Trp Glu Asn Gly Arg Glu
         65                  70                  75

TTA CCT GGC ATC AAT ACG ACA ATT TCC CAG GAT CCT GAA TCT GAA TTG       530
Leu Pro Gly Ile Asn Thr Thr Ile Ser Gln Asp Pro Glu Ser Glu Leu
     80                  85                  90

TAC ACC ATT AGT AGC CAA CTA GAT TTC AAT ACG ACT CGC AAC CAC ACC       578
Tyr Thr Ile Ser Ser Gln Leu Asp Phe Asn Thr Thr Arg Asn His Thr
 95                  100                 105                 110

ATT AAG TGT CTC ATT AAA TAT GGA GAT GCT CAC GTG TCA GAG GAC TTC       626
Ile Lys Cys Leu Ile Lys Tyr Gly Asp Ala His Val Ser Glu Asp Phe
                115                 120                 125

ACC TGG GAA AAA CCC CCA GAA GAC CCT CCT GAT AGC AAG AAC ACA CTT       674
Thr Trp Glu Lys Pro Pro Glu Asp Pro Pro Asp Ser Lys Asn Thr Leu
            130                 135                 140

GTG CTC TTT GGG GCA GGA TTC GGC GCA GTA ATA ACA GTC GTC GTC ATC       722
Val Leu Phe Gly Ala Gly Phe Gly Ala Val Ile Thr Val Val Val Ile
        145                 150                 155

GTT GTC ATC ATC AAA TGC TTC TGT AAG CAC AGA AGC TGT TTC AGA AGA       770
Val Val Ile Ile Lys Cys Phe Cys Lys His Arg Ser Cys Phe Arg Arg
    160                 165                 170
```

Cys Leu Ile Lys Tyr Gly Asp Ala His Val Ser Glu Asp Phe Thr Trp
115 120 125

Glu Lys Pro Pro Glu Asp Pro Pro Asp Ser Lys Asn Thr Leu Val Leu
130 135 140

Phe Gly Ala Gly Phe Gly Ala Val Ile Thr Val Val Val Ile Val Val
145 150 155 160

Ile Ile Lys Cys Phe Cys Lys His Arg Ser Cys Phe Arg Arg Asn Glu
165 170 175

Ala Ser Arg Glu Thr Asn Asn Ser Leu Thr Phe Gly Pro Glu Glu Ala
180 185 190

Leu Ala Glu Gln Thr Val Phe Leu
195 200

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1570 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 249..890

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAGTTTTATA CCTCAATAGA CTCTTACTAG TTTCTCTTTT TCAGGTTGTG AAACTCAACC 60

TTCAAAGACA CTCTGTTCCA TTTCTGTGGA CTAATAGGAT CATCTTTAGC ATCTGCCGGG 120

TGGATGCCAT CCAGGCTTCT TTTTCTACAT CTCTGTTTCT CGATTTTTGT GAGCCTAGGA 180

GGTGCCTAAG CTCCATTGGC TCTAGATTCC TGGCTTTCCC CATCATGTTC TCCAAAGCAT 240

CTGAAGCT ATG GCT TGC AAT TGT CAG TTG ATG CAG GAT ACA CCA CTC CTC 290
Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu
1 5 10

AAG TTT CCA TGT CCA AGG CTC AAT CTT CTC TTT GTG CTG CTG ATT CGT 338
Lys Phe Pro Cys Pro Arg Leu Asn Leu Leu Phe Val Leu Leu Ile Arg
15 20 25 30

CTT TCA CAA GTG TCT TCA GCT GAC TTC TCT ACC CCC AAC ATA ACT GAG 386
Leu Ser Gln Val Ser Ser Ala Asp Phe Ser Thr Pro Asn Ile Thr Glu
35 40 45

TCT GGA AAC CCA TCT GCA GAC ACT AAA AGG ATT ACC TGC TTT GCT TCC 434
Ser Gly Asn Pro Ser Ala Asp Thr Lys Arg Ile Thr Cys Phe Ala Ser
50 55 60

GTAGTGTAAT TTGTAGGAAA CATGTTGCTA TGTATACCCA TTTGAGGGTA ATAAAAATGT 1520

TGGTAATTTT CAGCCAGCAC TTTCCAGGTA TTTCCCTTTT TATCCTTCAT 1570

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu Lys Phe
1 5 10 15

Pro Cys Pro Arg Leu Ile Leu Leu Phe Val Leu Leu Ile Arg Leu Ser
20 25 30

Gln Val Ser Ser Ala Asp Phe Ser Thr Pro Asn Ile Thr Glu Ser Gly
35 40 45

Asn Pro Ser Ala Asp Thr Lys Arg Ile Thr Cys Phe Ala Ser Gly Gly
50 55 60

Phe Pro Lys Pro Arg Phe Ser Trp Leu Glu Asn Gly Arg Glu Leu Pro
65 70 75 80

Gly Ile Asn Thr Thr Ile Ser Gln Asp Pro Glu Ser Glu Leu Tyr Thr
85 90 95

Ile Ser Ser Gln Leu Asp Phe Asn Thr Thr Arg Asn His Thr Ile Lys
100 105 110

Cys Leu Ile Lys Tyr Gly Asp Ala His Val Ser Glu Asp Phe Thr Trp
115 120 125

Glu Lys Pro Pro Glu Asp Pro Pro Asp Ser Lys Asn Thr Leu Val Leu
130 135 140

Phe Gly Ala Gly Phe Gly Ala Val Ile Thr Val Val Val Ile Val Val
145 150 155 160

Ile Ile Lys Cys Phe Cys Lys His Gly Leu Ile Tyr His Leu Gln Leu
165 170 175

Thr Ser Ser Ala Lys Asp Phe Arg Asn Leu Ala Leu Pro Trp Leu Cys
180 185 190

Lys His Gly Ser Leu Gly Glu Ala Ser Ala Val Ile Cys Arg Ser Thr
195 200 205

Gln Thr Asn Glu Pro Gln
210

(2) INFORMATION FOR SEQ ID NO:12:

```
TTC AGT GAA CCT GAA ATA AAA CTG GCT CAG AAT GTA ACA GGA AAT TCT      661
Phe Ser Glu Pro Glu Ile Lys Leu Ala Gln Asn Val Thr Gly Asn Ser
                145                 150                 155

GGC ATA AAT TTG ACC TGC ACG TCT AAG CAA GGT CAC CCG AAA CCT AAG      709
Gly Ile Asn Leu Thr Cys Thr Ser Lys Gln Gly His Pro Lys Pro Lys
            160                 165                 170

AAG ATG TAT TTT CTG ATA ACT AAT TCA ACT AAT GAG TAT GGT GAT AAC      757
Lys Met Tyr Phe Leu Ile Thr Asn Ser Thr Asn Glu Tyr Gly Asp Asn
        175                 180                 185

ATG CAG ATA TCA CAA GAT AAT GTC ACA GAA CTG TTC AGT ATC TCC AAC      805
Met Gln Ile Ser Gln Asp Asn Val Thr Glu Leu Phe Ser Ile Ser Asn
    190                 195                 200

AGC CTC TCT CTT TCA TTC CCG GAT GGT GTG TGG CAT ATG ACC GTT GTG      853
Ser Leu Ser Leu Ser Phe Pro Asp Gly Val Trp His Met Thr Val Val
205                 210                 215                 220

TGT GTT CTG GAA ACG GAG TCA ATG AAG ATT TCC TCC AAA CCT CTC AAT      901
Cys Val Leu Glu Thr Glu Ser Met Lys Ile Ser Ser Lys Pro Leu Asn
                225                 230                 235

TTC ACT CAA GAG TTT CCA TCT CCT CAA ACG TAT TGG AAG GAG ATT ACA      949
Phe Thr Gln Glu Phe Pro Ser Pro Gln Thr Tyr Trp Lys Glu Ile Thr
            240                 245                 250

GCT TCA GTT ACT GTG GCC CTC CTC CTT GTG ATG CTG CTC ATC ATT GTA      997
Ala Ser Val Thr Val Ala Leu Leu Leu Val Met Leu Leu Ile Ile Val
        255                 260                 265

TGT CAC AAG AAG CCG AAT CAG CCT AGC AGG CCC AGC AAC ACA GCC TCT     1045
Cys His Lys Lys Pro Asn Gln Pro Ser Arg Pro Ser Asn Thr Ala Ser
    270                 275                 280

AAG TTA GAG CGG GAT AGT AAC GCT GAC AGA GAG ACT ATC AAC CTG AAG     1093
Lys Leu Glu Arg Asp Ser Asn Ala Asp Arg Glu Thr Ile Asn Leu Lys
285                 290                 295                 300

GAA CTT GAA CCC CAA ATT GCT TCA GCA AAA CCA AAT GCA GAG             1135
Glu Leu Glu Pro Gln Ile Ala Ser Ala Lys Pro Asn Ala Glu
                305                 310

TGAAGGCAGT GAGAGCCTGA GGAAAGAGTT AAAAATTGCT TTGCCTGAAA TAAGAAGTGC  1195

AGAGTTTCTC AGAATTCAAA AATGTTCTCA GCTGATTGGA ATTCTACAGT TGAATAATTA  1255

AAGAAC                                                             1261
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

Gln Ile Ala Ser Ala Lys Pro Asn Ala Glu
305 310

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 223 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 194..223

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGNCCCNAGA TTATTTCTCC CTGTATAAGG GACGCCCAGG AGGCCTGGGG AGCGGACAAG 60

GCTCCTTTTA CTTTTCTTCT TCTTCTATTT TTTTTACCTT CTATTTTTTT CTTCATGTTC 120

CTGTGATCTT CGGGAATGCT GCTGTGCTTG TGTGTGTGGT CCCTGAGCGC CGAGGTGGAG 180

AGGCACTGGT GAC ATG TAT GTC ATC AAG ACA TGT GCA ACC TGC 223
Met Tyr Val Ile Lys Thr Cys Ala Thr Cys
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Tyr Val Ile Lys Thr Cys Ala Thr Cys
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1716 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

```
CGC TTC TCT TGG TTG GAA AAT GGA AGA GAA TTA CCT GGC ATC AAT ACG      818
Arg Phe Ser Trp Leu Glu Asn Gly Arg Glu Leu Pro Gly Ile Asn Thr
175                 180                 185                 190

ACA ATT TCC CAG GAT CCT GAA TCT GAA TTG TAC ACC ATT AGT AGC CAA      866
Thr Ile Ser Gln Asp Pro Glu Ser Glu Leu Tyr Thr Ile Ser Ser Gln
                195                 200                 205

CTA GAT TTC AAT ACG ACT CGC AAC CAC ACC ATT AAG TGT CTC ATT AAA      914
Leu Asp Phe Asn Thr Thr Arg Asn His Thr Ile Lys Cys Leu Ile Lys
            210                 215                 220

TAT GGA GAT GCT CAC GTG TCA GAG GAC TTC ACC TGG GAA AAA CCC CCA      962
Tyr Gly Asp Ala His Val Ser Glu Asp Phe Thr Trp Glu Lys Pro Pro
        225                 230                 235

GAA GAC CCT CCT GAT AGC AAG AAC ACA CTT GTG CTC TTT GGG GCA GGA     1010
Glu Asp Pro Pro Asp Ser Lys Asn Thr Leu Val Leu Phe Gly Ala Gly
    240                 245                 250

TTC GGC GCA GTA ATA ACA GTC GTC GTC ATC GTT GTC ATC ATC AAA TGC     1058
Phe Gly Ala Val Ile Thr Val Val Val Ile Val Val Ile Ile Lys Cys
255                 260                 265                 270

TTC TGT AAG CAC AGA AGC TGT TTC AGA AGA AAT GAG GCA AGC AGA GAA     1106
Phe Cys Lys His Arg Ser Cys Phe Arg Arg Asn Glu Ala Ser Arg Glu
                275                 280                 285

ACA AAC AAC AGC CTT ACC TTC GGG CCT GAA GAA GCA TTA GCT GAA CAG     1154
Thr Asn Asn Ser Leu Thr Phe Gly Pro Glu Glu Ala Leu Ala Glu Gln
            290                 295                 300

ACC GTC TTC CTT TAGTTCTTCT CTGTCCATGT GGGATACATG GTATTATGTG         1206
Thr Val Phe Leu
        305

GCTCATGAGG TACAATCTTT CTTTCAGCAC CGTGCTAGCT GATCTTTCGG ACAACTTGAC   1266

ACAAGATAGA GTTAACTGGG AAGAGAAAGC CTTGAATGAG GATTTCTTTC CATCAGGAAG   1326

CTACGGGCAA GTTTGCTGGG CCTTTGATTG CTTGATGACT GAAGTGGAAA GGCTGAGCCC   1386

ACTGTGGGTG GTGCTAGCCC TGGGCAGGGG CAGGTGACCC TGGGTGGTAT AAGAAAAAGA   1446

GCTGTCACTA AAAGGAGAGG TGCCTAGTCT TACTGCAACT TGATATGTCA TGTTTGGTTG   1506

GTGTCTGTGG GAGGCCTGCC CTTTTCTGAA GAGAAGTGGT GGGAGAGTGG ATGGGGTGGG   1566

GGCAGAGGAA AAGTGGGGGA GAGGGCCTGG GAGGAGAGGA GGGAGGGGGA CGGGGTGGGG   1626

GTGGGGAAAA CTATGGTTGG GATGTAAAAA CGGATAATAA TATAAATATT AAATAAAAAG   1686

AGAGTATTGA GCAAAAAAAA AAAAAAAAAA                                    1716
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 306 amino acids

Lys His Arg Ser Cys Phe Arg Arg Asn Glu Ala Ser Arg Glu Thr Asn
275 280 285

Asn Ser Leu Thr Phe Gly Pro Glu Glu Ala Leu Ala Glu Gln Thr Val
290 295 300

Phe Leu
305

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1491 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 318..1181

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCAAAGAAAA AGTGATTTGT CATTGCTTTA TAGACTGTAA GAAGAGAACA TCTCAGAAGT 60

GGAGTCTTAC CCTGAAATCA AAGGATTTAA AGAAAAAGTG GAATTTTTCT TCAGCAAGCT 120

GTGAAACTAA ATCCACAACC TTTGGAGACC CAGGAACACC CTCCAATCTC TGTGTGTTTT 180

GTAAACATCA CTGGAGGGTC TTCTACGTGA GCAATTGGAT TGTCATCAGC CCTGCCTGTT 240

TTGCACCTGG GAAGTGCCCT GGTCTTACTT GGGTCCAAAT TGTTGGCTTT CACTTTTGAC 300

CCTAAGCATC TGAAGCC ATG GGC CAC ACA CGG AGG CAG GGA ACA TCA CCA 350
Met Gly His Thr Arg Arg Gln Gly Thr Ser Pro
1 5 10

TCC AAG TGT CCA TAC CTG AAT TTC TTT CAG CTC TTG GTG CTG GCT GGT 398
Ser Lys Cys Pro Tyr Leu Asn Phe Phe Gln Leu Leu Val Leu Ala Gly
15 20 25

CTT TCT CAC TTC TGT TCA GGT GTT ATC CAC GTG ACC AAG GAA GTG AAA 446
Leu Ser His Phe Cys Ser Gly Val Ile His Val Thr Lys Glu Val Lys
30 35 40

GAA GTG GCA ACG CTG TCC TGT GGT CAC AAT GTT TCT GTT GAA GAG CTG 494
Glu Val Ala Thr Leu Ser Cys Gly His Asn Val Ser Val Glu Glu Leu
45 50 55

GCA CAA ACT CGC ATC TAC TGG CAA AAG GAG AAG AAA ATG GTG CTG ACT 542
Ala Gln Thr Arg Ile Tyr Trp Gln Lys Glu Lys Lys Met Val Leu Thr
60 65 70 75

AAGCTGAACA GTTACAAGAT GGCTGGCATC CCTCTCCTTT CTCCCCATAT GCAATTTGCT 1401

TAATGTAACC TCTTCTTTTG CCATGTTTCC ATTCTGCCAT CTTGAATTGT CTTGTCAGCC 1461

AATTCATTAT CTATTAAACA CTAATTTGAG 1491

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 288 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Gly His Thr Arg Arg Gln Gly Thr Ser Pro Ser Lys Cys Pro Tyr
1 5 10 15

Leu Asn Phe Phe Gln Leu Leu Val Leu Ala Gly Leu Ser His Phe Cys
20 25 30

Ser Gly Val Ile His Val Thr Lys Glu Val Lys Glu Val Ala Thr Leu
35 40 45

Ser Cys Gly His Asn Val Ser Val Glu Glu Leu Ala Gln Thr Arg Ile
50 55 60

Tyr Trp Gln Lys Glu Lys Lys Met Val Leu Thr Met Met Ser Gly Asp
65 70 75 80

Met Asn Ile Trp Pro Glu Tyr Lys Asn Arg Thr Ile Phe Asp Ile Thr
85 90 95

Asn Asn Leu Ser Ile Val Ile Leu Ala Leu Arg Pro Ser Asp Glu Gly
100 105 110

Thr Tyr Glu Cys Val Val Leu Lys Tyr Glu Lys Asp Ala Phe Lys Arg
115 120 125

Glu His Leu Ala Glu Val Thr Leu Ser Val Lys Ala Asp Phe Pro Thr
130 135 140

Pro Ser Ile Ser Asp Phe Glu Ile Pro Thr Ser Asn Ile Arg Arg Ile
145 150 155 160

Ile Cys Ser Thr Ser Gly Gly Phe Pro Glu Pro His Leu Ser Trp Leu
165 170 175

Glu Asn Gly Glu Glu Leu Asn Ala Ile Asn Thr Thr Val Ser Gln Asp
180 185 190

Pro Glu Thr Glu Leu Tyr Ala Val Ser Ser Lys Leu Asp Phe Asn Met
195 200 205

```
ACG AGC TTT GAC AGG AAC AAC TGG ACT CTA CGA CTT CAC AAT GTT CAG      401
Thr Ser Phe Asp Arg Asn Asn Trp Thr Leu Arg Leu His Asn Val Gln
                 90                  95                 100

ATC AAG GAC ATG GGC TCG TAT GAT TGT TTT ATA CAA AAA AAG CCA CCC      449
Ile Lys Asp Met Gly Ser Tyr Asp Cys Phe Ile Gln Lys Lys Pro Pro
            105                 110                 115

ACA GGA TCA ATT ATC CTC CAA CAG ACA TTA ACA GAA CTG TCA GTG ATC      497
Thr Gly Ser Ile Ile Leu Gln Gln Thr Leu Thr Glu Leu Ser Val Ile
        120                 125                 130

GCC AAC TTC AGT GAA CCT GAA ATA AAA CTG GCT CAG AAT GTA ACA GGA      545
Ala Asn Phe Ser Glu Pro Glu Ile Lys Leu Ala Gln Asn Val Thr Gly
    135                 140                 145

AAT TCT GGC ATA AAT TTG ACC TGC ACG TCT AAG CAA GGT CAC CCG AAA      593
Asn Ser Gly Ile Asn Leu Thr Cys Thr Ser Lys Gln Gly His Pro Lys
150                 155                 160                 165

CCT AAG AAG ATG TAT TTT CTG ATA ACT AAT TCA ACT AAT GAG TAT GGT      641
Pro Lys Lys Met Tyr Phe Leu Ile Thr Asn Ser Thr Asn Glu Tyr Gly
                170                 175                 180

GAT AAC ATG CAG ATA TCA CAA GAT AAT GTC ACA GAA CTG TTC AGT ATC      689
Asp Asn Met Gln Ile Ser Gln Asp Asn Val Thr Glu Leu Phe Ser Ile
            185                 190                 195

TCC AAC AGC CTC TCT CTT TCA TTC CCG GAT GGT GTG TGG CAT ATG ACC      737
Ser Asn Ser Leu Ser Leu Ser Phe Pro Asp Gly Val Trp His Met Thr
        200                 205                 210

GTT GTG TGT GTT CTG GAA ACG GAG TCA ATG AAG ATT TCC TCC AAA CCT      785
Val Val Cys Val Leu Glu Thr Glu Ser Met Lys Ile Ser Ser Lys Pro
    215                 220                 225

CTC AAT TTC ACT CAA GAG TTT CCA TCT CCT CAA ACG TAT TGG AAG GAG      833
Leu Asn Phe Thr Gln Glu Phe Pro Ser Pro Gln Thr Tyr Trp Lys Glu
230                 235                 240                 245

ATT ACA GCT TCA GTT ACT GTG GCC CTC CTC CTT GTG ATG CTG CTC ATC      881
Ile Thr Ala Ser Val Thr Val Ala Leu Leu Leu Val Met Leu Leu Ile
                250                 255                 260

ATT GTA TGT CAC AAG AAG CCG AAT CAG CCT AGC AGG CCC AGC AAC ACA      929
Ile Val Cys His Lys Lys Pro Asn Gln Pro Ser Arg Pro Ser Asn Thr
            265                 270                 275

GCC TCT AAG TTA GAG CGG GAT AGT AAC GCT GAC AGA GAG ACT ATC AAC      977
Ala Ser Lys Leu Glu Arg Asp Ser Asn Ala Asp Arg Glu Thr Ile Asn
        280                 285                 290

CTG AAG GAA CTT GAA CCC CAA ATT GCT TCA GCA AAA CCA AAT GCA GAG     1025
Leu Lys Glu Leu Glu Pro Gln Ile Ala Ser Ala Lys Pro Asn Ala Glu
    295                 300                 305

TGAAGGCAGT GAGAGCCTGA GGAAAGAGTT AAAAATTGCT TTGCCTGAAA TAAGAAGTGC   1085
```

```
Ile Ser Ser Lys Pro Leu Asn Phe Thr Gln Glu Phe Pro Ser Pro Gln
225                 230                 235                 240

Thr Tyr Trp Lys Glu Ile Thr Ala Ser Val Thr Val Ala Leu Leu Leu
                245                 250                 255

Val Met Leu Leu Ile Ile Val Cys His Lys Lys Pro Asn Gln Pro Ser
            260                 265                 270

Arg Pro Ser Asn Thr Ala Ser Lys Leu Glu Arg Asp Ser Asn Ala Asp
        275                 280                 285

Arg Glu Thr Ile Asn Leu Lys Glu Leu Glu Pro Gln Ile Ala Ser Ala
    290                 295                 300

Lys Pro Asn Ala Glu
305
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1120 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 107..1093

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
CACAGGGTGA AAGCTTTGCT TCTCTGCTGC TGTAACAGGG ACTAGCACAG ACACACGGAT      60

GAGTGGGGTC ATTTCCAGAT ATTAGGTCAC AGCAGAAGCA GCCAAA ATG GAT CCC        115
                                                    Met Asp Pro
                                                      1

CAG TGC ACT ATG GGA CTG AGT AAC ATT CTC TTT GTG ATG GCC TTC CTG       163
Gln Cys Thr Met Gly Leu Ser Asn Ile Leu Phe Val Met Ala Phe Leu
  5                  10                  15

CTC TCT GGT GCT GCT CCT CTG AAG ATT CAA GCT TAT TTC AAT GAG ACT       211
Leu Ser Gly Ala Ala Pro Leu Lys Ile Gln Ala Tyr Phe Asn Glu Thr
 20                  25                  30                  35

GCA GAC CTG CCA TGC CAA TTT GCA AAC TCT CAA AAC CAA AGC CTG AGT       259
Ala Asp Leu Pro Cys Gln Phe Ala Asn Ser Gln Asn Gln Ser Leu Ser
                 40                  45                  50

GAG CTA GTA GTA TTT TGG CAG GAC CAG GAA AAC TTG GTT CTG AAT GAG       307
Glu Leu Val Val Phe Trp Gln Asp Gln Glu Asn Leu Val Leu Asn Glu
             55                  60                  65
```

```
GAT GAA GCC CAG CGT GTT TTT AAA AGT TCG AAG ACA TCT TCA TGC GAC      1075
Asp Glu Ala Gln Arg Val Phe Lys Ser Ser Lys Thr Ser Ser Cys Asp
        310                 315                 320

AAA AGT GAT ACA TGT TTT TAATTAAAGA GTAAAGCCCA AAAAAAA               1120
Lys Ser Asp Thr Cys Phe
    325
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 329 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Met Asp Pro Gln Cys Thr Met Gly Leu Ser Asn Ile Leu Phe Val Met
 1               5                  10                  15

Ala Phe Leu Leu Ser Gly Ala Ala Pro Leu Lys Ile Gln Ala Tyr Phe
            20                  25                  30

Asn Glu Thr Ala Asp Leu Pro Cys Gln Phe Ala Asn Ser Gln Asn Gln
        35                  40                  45

Ser Leu Ser Glu Leu Val Val Phe Trp Gln Asp Gln Glu Asn Leu Val
    50                  55                  60

Leu Asn Glu Val Tyr Leu Gly Lys Glu Lys Phe Asp Ser Val His Ser
65                  70                  75                  80

Lys Tyr Met Gly Arg Thr Ser Phe Asp Ser Asp Ser Trp Thr Leu Arg
                85                  90                  95

Leu His Asn Leu Gln Ile Lys Asp Lys Gly Leu Tyr Gln Cys Ile Ile
            100                 105                 110

His His Lys Lys Pro Thr Gly Met Ile Arg Ile His Gln Met Asn Ser
        115                 120                 125

Glu Leu Ser Val Leu Ala Asn Phe Ser Gln Pro Glu Ile Val Pro Ile
    130                 135                 140

Ser Asn Ile Thr Glu Asn Val Tyr Ile Asn Leu Thr Cys Ser Ser Ile
145                 150                 155                 160

His Gly Tyr Pro Glu Pro Lys Lys Met Ser Val Leu Leu Arg Thr Lys
                165                 170                 175

Asn Ser Thr Ile Glu Tyr Asp Gly Ile Met Gln Lys Ser Gln Asp Asn
            180                 185                 190

Val Thr Glu Leu Tyr Asp Val Ser Ile Ser Leu Ser Val Ser Phe Pro
        195                 200                 205
```

```
CAA TTT GCA AAC TCT CAA AAC CAA AGC CTG AGT GAG CTA GTA GTA TTT      315
Gln Phe Ala Asn Ser Gln Asn Gln Ser Leu Ser Glu Leu Val Val Phe
                45                  50                  55

TGG CAG GAC CAG GAA AAC TTG GTT CTG AAT GAG GTA TAC TTA GGC AAA      363
Trp Gln Asp Gln Glu Asn Leu Val Leu Asn Glu Val Tyr Leu Gly Lys
            60                  65                  70

GAG AAA TTT GAC AGT GTT CAT TCC AAG TAT ATG GGC CGC ACA AGT TTT      411
Glu Lys Phe Asp Ser Val His Ser Lys Tyr Met Gly Arg Thr Ser Phe
        75                  80                  85

GAT TCG GAC AGT TGG ACC CTG AGA CTT CAC AAT CTT CAG ATC AAG GAC      459
Asp Ser Asp Ser Trp Thr Leu Arg Leu His Asn Leu Gln Ile Lys Asp
    90                  95                  100

AAG GGC TTG TAT CAA TGT ATC ATC CAT CAC AAA AAG CCC ACA GGA ATG      507
Lys Gly Leu Tyr Gln Cys Ile Ile His His Lys Lys Pro Thr Gly Met
105                 110                 115                 120

ATT CGC ATC CAC CAG ATG AAT TCT GAA CTG TCA GTG CTT GCT AAC TTC      555
Ile Arg Ile His Gln Met Asn Ser Glu Leu Ser Val Leu Ala Asn Phe
                125                 130                 135

AGT CAA CCT GAA ATA GTA CCA ATT TCT AAT ATA ACA GAA AAT GTG TAC      603
Ser Gln Pro Glu Ile Val Pro Ile Ser Asn Ile Thr Glu Asn Val Tyr
            140                 145                 150

ATA AAT TTG ACC TGC TCA TCT ATA CAC GGT TAC CCA GAA CCT AAG AAG      651
Ile Asn Leu Thr Cys Ser Ser Ile His Gly Tyr Pro Glu Pro Lys Lys
        155                 160                 165

ATG AGT GTT TTG CTA AGA ACC AAG AAT TCA ACT ATC GAG TAT GAT GGT      699
Met Ser Val Leu Leu Arg Thr Lys Asn Ser Thr Ile Glu Tyr Asp Gly
    170                 175                 180

ATT ATG CAG AAA TCT CAA GAT AAT GTC ACA GAA CTG TAC GAC GTT TCC      747
Ile Met Gln Lys Ser Gln Asp Asn Val Thr Glu Leu Tyr Asp Val Ser
185                 190                 195                 200

ATC AGC TTG TCT GTT TCA TTC CCT GAT GTT ACG AGC AAT ATG ACC ATC      795
Ile Ser Leu Ser Val Ser Phe Pro Asp Val Thr Ser Asn Met Thr Ile
                205                 210                 215

TTC TGT ATT CTG GAA ACT GAC AAG ACG CGG CTT TTA TCT TCA CCT TTC      843
Phe Cys Ile Leu Glu Thr Asp Lys Thr Arg Leu Leu Ser Ser Pro Phe
            220                 225                 230

TCT ATA GAG CTT GAG GAC CCT CAG CCT CCC CCA GAC CAC ATT CCT TGG      891
Ser Ile Glu Leu Glu Asp Pro Gln Pro Pro Pro Asp His Ile Pro Trp
        235                 240                 245

ATT ACA GCT GTA CTT CCA ACA GTT ATT ATA TGT GTG ATG GTT TTC TGT      939
Ile Thr Ala Val Leu Pro Thr Val Ile Ile Cys Val Met Val Phe Cys
    250                 255                 260

CTA ATT CTA TGG AAA TGG AAG AAG AAG AAG CGG CCT CGC AAC TCT TAT      987
Leu Ile Leu Trp Lys Trp Lys Lys Lys Lys Arg Pro Arg Asn Ser Tyr
265                 270                 275                 280
```

GATGTAAAAA CGGATAATAA TATAAATATT AAATAAAAAG AGAGTATTGA GCA 629

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Arg Ser Cys Phe Arg Arg Asn Glu Ala Ser Arg Glu Thr Asn Asn Ser
1 5 10 15

Leu Thr Phe Gly Pro Glu Glu Ala Leu Ala Glu Gln Thr Val Phe Leu
20 25 30

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 379 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..69

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGC TTT GCC CCA AGA TGC AGA GAG AGA AGG AGG AAT GAG AGA TTG AGA 48
Cys Phe Ala Pro Arg Cys Arg Glu Arg Arg Arg Asn Glu Arg Leu Arg
1 5 10 15

AGG GAA AGT GTA CGC CCT GTA TAACAGTGTC CGCAGAAGCA AGGGGCTGAA 99
Arg Glu Ser Val Arg Pro Val
20

AAGATCTGAA GGTAGCCTCC GTCATCTCTT CTGGGATACA TGGATCGTGG GGATCATGAG 159

GCATTCTTCC CTTAACAAAT TTAAGCTGTT TTACCCACTA CCTCACCTTC TTAAAAACCT 219

CTTTCAGATT AAGCTGAACA GTTACAAGAT GGCTGGCATC CCTCTCCTTT CTCCCCATAT 279

GCAATTTGCT TAATGTAACC TCTTCTTTTG CCATGTTTCC ATTCTGCCAT CTTGAATTGT 339

CTTGTCAGCC AATTCATTAT CTATTAAACA CTAATTTGAG 379

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
His Lys Lys Pro Asn Gln Pro Ser Arg Pro Ser Asn Thr Ala Ser Lys
 1               5                  10                  15

Leu Glu Arg Asp Ser Asn Ala Asp Arg Glu Thr Ile Asn Leu Lys Glu
            20                  25                  30

Leu Glu Pro Gln Ile Ala Ser Ala Lys Pro Asn Ala Glu
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 210 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..183

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
AAA TGG AAG AAG AAG AAG CGG CCT CGC AAC TCT TAT AAA TGT GGA ACC      48
Lys Trp Lys Lys Lys Lys Arg Pro Arg Asn Ser Tyr Lys Cys Gly Thr
 1               5                  10                  15

AAC ACA ATG GAG AGG GAA GAG AGT GAA CAG ACC AAG AAA AGA GAA AAA      96
Asn Thr Met Glu Arg Glu Glu Ser Glu Gln Thr Lys Lys Arg Glu Lys
            20                  25                  30

ATC CAT ATA CCT GAA AGA TCT GAT GAA GCC CAG CGT GTT TTT AAA AGT     144
Ile His Ile Pro Glu Arg Ser Asp Glu Ala Gln Arg Val Phe Lys Ser
        35                  40                  45

TCG AAG ACA TCT TCA TGC GAC AAA AGT GAT ACA TGT TTT TAATTAAAGA     193
Ser Lys Thr Ser Ser Cys Asp Lys Ser Asp Thr Cys Phe
    50                  55                  60

GTAAAGCCCA AAAAAAA                                                 210
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu Lys Phe
1 5 10 15

Pro Cys Pro Arg Leu Ile Leu Leu Phe Val Leu Leu Ile Arg Leu Ser
20 25 30

Gln Val Ser Ser Asp
35

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 318..416

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCAAAGAAAA AGTGATTTGT CATTGCTTTA TAGACTGTAA GAAGAGAACA TCTCAGAAGT 60

GGAGTCTTAC CCTGAAATCA AAGGATTTAA AGAAAAAGTG GAATTTTTCT TCAGCAAGCT 120

GTGAAACTAA ATCCACAACC TTTGGAGACC CAGGAACACC CTCCAATCTC TGTGTGTTTT 180

GTAAACATCA CTGGAGGGTC TTCTACGTGA GCAATTGGAT TGTCATCAGC CCTGCCTGTT 240

TTGCACCTGG GAAGTGCCCT GGTCTTACTT GGGTCCAAAT TGTTGGCTTT CACTTTTGAC 300

CCTAAGCATC TGAAGCC ATG GGC CAC ACA CGG AGG CAG GGA ACA TCA CCA 350
Met Gly His Thr Arg Arg Gln Gly Thr Ser Pro
1 5 10

TCC AAG TGT CCA TAC CTG AAT TTC TTT CAG CTC TTG GTG CTG GCT GGT 398
Ser Lys Cys Pro Tyr Leu Asn Phe Phe Gln Leu Leu Val Leu Ala Gly
15 20 25

CTT TCT CAC TTC TGT TCA 416
Leu Ser His Phe Cys Ser
30

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 107..124

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CACAGGGTGA AAGCTTTGCT TCTCTGCTGC TGTAACAGGG ACTAGCACAG ACACACGGAT 60

GAGTGGGGTC ATTTCCAGAT ATTAGGTCAC AGCAGAAGCA GCCAAA ATG GAT CCC 115
Met Asp Pro
1

CAG TGC ACT 124
Gln Cys Thr
5

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Met Asp Pro Gln Cys Thr
1 5

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 195 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 148..195

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGGAGCCTTA GGAGGTACGG GGAGCTCGCA AATACTCCTT TTGGTTTATT CTTACCACCT 60

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

AGGTTAAGAG TGGTAGAGCC A 21

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

AATACCATGT ATCCCACATG G 21

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CTGAAGCTAT GGCTTGCAAT T 21

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

TGGCTTCTCT TTCCTTACCT T 21

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

ACTGACTTGG ACAGTTGTTC A 21

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TTTGATGGAC AACTTTACTA 20

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CAGCTCACTC AGGCTTATGT 20

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

AAACAGCATC TGAGATCAGC A 21

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CTGAGATCAG CAAGACTGTC 20

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1417 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 249..884

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

```
GAGTTTTATA CCTCAATAGA CTCTTACTAG TTTCTCTTTT TCAGGTTGTG AAACTCAACC      60

TTCAAAGACA CTCTGTTCCA TTTCTGTGGA CTAATAGGAT CATCTTTAGC ATCTGCCGGG     120

TGGATGCCAT CCAGGCTTCT TTTTCTACAT CTCTGTTTCT CGATTTTTGT GAGCCTAGGA     180

GGTGCCTAAG CTCCATTGGC TCTAGATTCC TGGCTTTCCC CATCATGTTC TCCAAAGCAT     240

CTGAAGCT ATG GCT TGC AAT TGT CAG TTG ATG CAG GAT ACA CCA CTC CTC       290
         Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu
           1               5                  10

AAG TTT CCA TGT CCA AGG CTC AAT CTT CTC TTT GTG CTG CTG ATT CGT       338
Lys Phe Pro Cys Pro Arg Leu Asn Leu Leu Phe Val Leu Leu Ile Arg
 15                  20                  25                  30

CTT TCA CAA GTG TCT TCA GAT GTT GAT GAA CAA CTG TCC AAG TCA GTG       386
Leu Ser Gln Val Ser Ser Asp Val Asp Glu Gln Leu Ser Lys Ser Val
                 35                  40                  45

AAA GAT AAG GTA TTG CTG CCT TGC CGT TAC AAC TCT CCT CAT GAA GAT       434
Lys Asp Lys Val Leu Leu Pro Cys Arg Tyr Asn Ser Pro His Glu Asp
             50                  55                  60

GAG TCT GAA GAC CGA ATC TAC TGG CAA AAA CAT GAC AAA GTG GTG CTG       482
Glu Ser Glu Asp Arg Ile Tyr Trp Gln Lys His Asp Lys Val Val Leu
         65                  70                  75

TCT GTC ATT GCT GGG AAA CTA AAA GTG TGG CCC GAG TAT AAG AAC CGG       530
Ser Val Ile Ala Gly Lys Leu Lys Val Trp Pro Glu Tyr Lys Asn Arg
     80                  85                  90

ACT TTA TAT GAC AAC ACT ACC TAC TCT CTT ATC ATC CTG GGC CTG GTC       578
Thr Leu Tyr Asp Asn Thr Thr Tyr Ser Leu Ile Ile Leu Gly Leu Val
 95                  100                 105                 110

CTT TCA GAC CGG GGC ACA TAC AGC TGT GTC GTT CAA AAG AAG GAA AGA       626
Leu Ser Asp Arg Gly Thr Tyr Ser Cys Val Val Gln Lys Lys Glu Arg
                115                 120                 125

GGA ACG TAT GAA GTT AAA CAC TTG GCT TTA GTA AAG TTG TCC ATC AAA       674
Gly Thr Tyr Glu Val Lys His Leu Ala Leu Val Lys Leu Ser Ile Lys
            130                 135                 140
```

Lys Val Leu Leu Pro Cys Arg Tyr Asn Ser Pro His Glu Asp Glu Ser
50 55 60

Glu Asp Arg Ile Tyr Trp Gln Lys His Asp Lys Val Val Leu Ser Val
65 70 75 80

Ile Ala Gly Lys Leu Lys Val Trp Pro Glu Tyr Lys Asn Arg Thr Leu
85 90 95

Tyr Asp Asn Thr Thr Tyr Ser Leu Ile Ile Leu Gly Leu Val Leu Ser
100 105 110

Asp Arg Gly Thr Tyr Ser Cys Val Val Gln Lys Lys Glu Arg Gly Thr
115 120 125

Tyr Glu Val Lys His Leu Ala Leu Val Lys Leu Ser Ile Lys Pro Pro
130 135 140

Glu Asp Pro Pro Asp Ser Lys Asn Thr Leu Val Leu Phe Gly Ala Gly
145 150 155 160

Phe Gly Ala Val Ile Thr Val Val Val Ile Val Val Ile Ile Lys Cys
165 170 175

Phe Cys Lys His Arg Ser Cys Phe Arg Arg Asn Glu Ala Ser Arg Glu
180 185 190

Thr Asn Asn Ser Leu Thr Phe Gly Pro Glu Glu Ala Leu Ala Glu Gln
195 200 205

Thr Val Phe Leu
210

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1606 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 249..926

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAGTTTTATA CCTCAATAGA CTCTTACTAG TTTCTCTTTT TCAGGTTGTG AAACTCAACC 60

TTCAAAGACA CTCTGTTCCA TTTCTGTGGA CTAATAGGAT CATCTTTAGC ATCTGCCGGG 120

TGGATGCCAT CCAGGCTTCT TTTTCTACAT CTCTGTTTCT CGATTTTTGT GAGCCTAGGA 180

GGTGCCTAAG CTCCATTGGC TCTAGATTCC TGGCTTTCCC CATCATGTTC TCCAAAGCAT 240

AAT GAA CCA CAG TAGTTCTGCT GTTTCTGAGG ACGTAGTTTA GAGACTGAAT 966
Asn Glu Pro Gln
225

TCTTTGGAAA GGACATAGGG ACAGTTTGCA CATTTGCTTG CACATCACAC ACACACACAC 1026

ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC TCTCTCTCTC 1086

TCTCTCTCTC GATACCTTAG GATAGGGTTC TACCCTGTTG CTCAGTGACA AAGAATCACT 1146

CTGTGGCGGA GGCAGGCTTC AAGCTTGCAG CAATCCTCCT GCACCAGTTT CCTGAGTGCC 1206

AGACTTCCAG GTGTAAGCTA TGGCACTTAG CAGAACACTA GCTGAATCAA TGAAGACACT 1266

GAGGTTCCAA GAGGGAACCT GAATTATGAA GGTGAGTCAG AATCCAGATT TCCTGGCTCT 1326

ACCACTCTTA ACCTGTATCT GTTAGACCCC AAGCTCTGAG CTCATAGACA AGCTAATTTA 1386

AAATGCTTTT TAATAAGCAG AAGGCTCAGT TAGTACGGGG TTCAGGATAC TGCTTACTGG 1446

CAATATTTGA CTAGCCTCTA TTTTGTTTGT TTTTTAAAGG CCTACTGACT GTAGTGTAAT 1506

TTGTAGGAAA CATGTTGCTA TGTATACCCA TTTGAGGGTA ATAAAAATGT TGGTAATTTT 1566

CAGCCAGCAC TTTCCAGGTA TTTCCCTTTT TATCCTTCAT 1606

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 226 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Met Ala Cys Asn Cys Gln Leu Met Gln Asp Thr Pro Leu Leu Lys Phe
1 5 10 15

Pro Cys Pro Arg Leu Ile Leu Leu Phe Val Leu Leu Ile Arg Leu Ser
20 25 30

Gln Val Ser Ser Asp Val Asp Glu Gln Leu Ser Lys Ser Val Lys Asp
35 40 45

Lys Val Leu Leu Pro Cys Arg Tyr Asn Ser Pro His Glu Asp Glu Ser
50 55 60

Glu Asp Arg Ile Tyr Trp Gln Lys His Asp Lys Val Val Leu Ser Val
65 70 75 80

Ile Ala Gly Lys Leu Lys Val Trp Pro Glu Tyr Lys Asn Arg Thr Leu
85 90 95

Tyr Asp Asn Thr Thr Tyr Ser Leu Ile Ile Leu Gly Leu Val Leu Ser
100 105 110

CLAIMS

1. An isolated nucleic acid encoding a protein which binds CD28 or CTLA4 comprising a contiguous nucleotide sequence derived from at least one T cell costimulatory molecule gene, the nucleotide sequence represented by a formula A-B-C-D-E, wherein

A comprises a nucleotide sequence of at least one first exon of a T cell costimulatory molecule gene, wherein the at least one first exon encodes a signal peptide domain,

B comprises a nucleotide sequence of at least one second exon of a T cell costimulatory molecule gene, wherein the at least one second exon encodes an immunoglobulin variable region-like domain,

C comprises a nucleotide sequence of at least one third exon of a T cell costimulatory molecule gene, wherein the at least one third exon encodes an immunoglobulin constant region-like domain,

D comprises a nucleotide sequence of at least one fourth exon of a T cell costimulatory molecule gene, wherein the at least one fourth exon encodes a transmembrane domain, and

E comprises a nucleotide sequence of at least one fifth exon of a T cell costimulatory molecule gene, wherein the at least one fifth exon encodes a cytoplasmic domain,

with the proviso that E does not comprise a nucleotide sequence selected from a group consisting of SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29 and SEQ ID NO:31.

2. The isolated nucleic acid of claim 1 which is a cDNA.

3. The isolated nucleic acid of claim 2 which comprises a coding region of the cDNA.

4. The isolated nucleic acid of claim 1, wherein the nucleotide sequence is derived from a T cell costimulatory molecule gene encoding B7-1.

5. The isolated nucleic acid of claim 4, wherein B7-1 is murine.

6. The isolated nucleic acid of claim 4, wherein B7-1 is human.

7. The isolated nucleic acid of claim 5, wherein E comprises a nucleotide sequence shown in SEQ ID NO:4.

18. An isolated protein which binds to CD28 or CTLA4 having an amino acid sequence derived from amino acid sequences encoded by at least one T cell costimulatory molecule gene, the protein comprising a contiguous amino acid sequence represented by a formula A-B-C-D-E, wherein

A, which may or may not be present, comprises an amino acid sequence of a signal peptide domain encoded by at least one exon of a T cell costimulatory molecule gene,

B comprises an amino acid sequence of an immunoglobulin variable region-like domain encoded by at least one exon of a T cell costimulatory molecule gene,

C comprises an amino acid sequence of an immunoglobulin constant region-like domain encoded by at least one exon of aT cell costimulatory molecule gene,

D comprises an amino acid sequence of a transmembrane domain encoded by at least one exon of a T cell costimulatory molecule gene, and

E comprises an amino acid sequence of a cytoplasmic domain encoded by at least one exon of a T cell costimulatory molecule gene,

with the proviso that E not comprise an amino acid sequence selected from the group consisting of SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30 and SEQ ID NO:32.

19. The isolated protein of claim 18 which is B7-1.

20. The isolated protein of claim 19 which is murine.

21. The isolated protein of claim 19 which is human.

22. The isolated protein of claim 20, wherein E comprises an amino acid sequence shown in SEQ ID NO:5.

23. An isolated protein which binds CD28 or CTLA4 and is encoded by a T cell costimulatory molecule gene having

at least one first exon encoding a first cytoplasmic domain comprising an amino acid sequence selected from the group consisting of an amino acid sequence of SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:32, and

at least one second exon encoding a second cytoplasmic domain,

wherein the T cell costimulatory molecule comprises the second cytoplasmic domain.

24. The isolated protein of claim 23 which does not comprise the first cytoplasmic domain.

E, which may or may not be present, comprises a nucleotide sequence of at least one fifth exon of a T cell costimulatory molecule gene, wherein the at least one fifth exon encodes a cytoplasmic domain,

with the proviso that A does not comprise a nucleotide sequence selected from a group consisting of SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39 and SEQ ID NO:41.

34. The isolated nucleic acid of claim 33 which is a cDNA.

35. The isolated nucleic acid of claim 34 which comprises a coding region of the cDNA.

36. The isolated nucleic acid of claim 33, wherein the nucleotide sequence is derived from a T cell costimulatory molecule gene encoding B7-2.

37. The isolated nucleic acid of claim 36, wherein B7-2 is murine.

38. The isolated nucleic acid of claim 36, wherein B7-2 is human.

39. The isolated nucleic acid of claim 37, wherein A comprises a nucleotide sequence shown in SEQ ID NO:14.

40. An isolated nucleic acid encoding a protein which binds CD28 or CTLA4 and is encoded by a T cell costimulatory molecule gene having

at least one first exon encoding a first signal peptide domain comprising a nucleotide sequence selected from the group consisting of a nucleotide sequence of SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 SEQ ID NO:39 and SEQ ID NO:41, and

at least one second exon encoding a second signal peptide domain,

wherein the isolated nucleic acid comprises a nucleotide sequence encoding the second signal peptide domain.

41. The isolated nucleic acid of claim 40 which comprises a coding region of a cDNA.

42. The isolated nucleic acid of claim 40 which does not comprise a nucleotide sequence encoding the first signal peptide domain.

51. The isolated protein of claim 49 which is human.

52. The isolated protein of claim 50, wherein A comprises an amino acid sequence shown in SEQ ID NO: 15.

53. An isolated protein which binds CD28 or CTLA4 and is encoded by a T cell costimulatory molecule gene having

at least one first exon encoding a first signal peptide domain comprising an amino acid sequence selected from the group consisting of an amino acid sequence of SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40 and SEQ ID NO:42, and

at least one second exon encoding a second signal peptide domain,

wherein the T cell costimulatory molecule comprises the second signal peptide domain.

54. The isolated protein of claim 53 which does not comprise the first signal peptide domain.

55. The isolated protein of claim 53 which is B7-2.

56. The isolated protein of claim 55 which is murine.

57. The isolated protein of claim 55 which is human.

58. An isolated protein which binds CD28 or CTLA4 comprising an amino acid sequence shown in SEQ ID NO:13.

59. An isolated signal peptide domain polypeptide derived from a protein which binds CD28 or CTLA4, the polypeptide comprising an amino acid sequence shown in SEQ ID NO:15.

60. A recombinant expression vector comprising the nucleic acid molecule of claim 46.

61. A host cell which contains the recombinant expression vector of claim 60.

62. An antibody which binds to the polypeptide of claim 59.

68. The isolated protein of claim 66 comprising an amino acid sequence shown in SEQ ID NO:11.

69. An isolated nucleic acid encoding a protein comprising a contiguous nucleotide sequence derived from at least one T cell costimulatory molecule gene, the nucleotide sequence represented by a formula A-B-C-D, wherein

A comprises a nucleotide sequence of at least one first exon of a T cell costimulatory molecule gene, wherein the at least one first exon encodes a signal peptide domain,

B comprises a nucleotide sequence of at least one second exon of a T cell costimulatory molecule gene, wherein the at least one second exon encodes an immunoglobulin variable region-like domain,

C comprises a nucleotide sequence of at least one third exon of a T cell costimulatory molecule gene, wherein the at least one third exon encodes a transmembrane domain, and

D comprises a nucleotide sequence of at least one fourth exon of a T cell costimulatory molecule gene, wherein the at least one fourth exon encodes a cytoplasmic domain.

70. The isolated nucleic acid of claim 69 comprising a nucleotide sequence shown in SEQ ID NO:62.

71. The isolated nucleic acid of claim 69 comprising a nucleotide sequence shown in SEQ ID NO:64.

72. An isolated protein having an amino acid sequence derived from amino acid sequences encoded by at least one T cell costimulatory molecule gene, the protein comprising a contiguous amino acid sequence represented by a formula A-B-C-D, wherein

A, which may or may not be present, comprises an amino acid sequence of a signal peptide domain encoded by at least one exon of a T cell costimulatory molecule gene,

B comprises an amino acid sequence of an immunoglobulin variable region-like domain encoded by at least one exon of a T cell costimulatory molecule gene, and

C comprises an amino acid sequence of a transmembrane domain encoded by at least one exon of a T cell costimulatory molecule gene, and

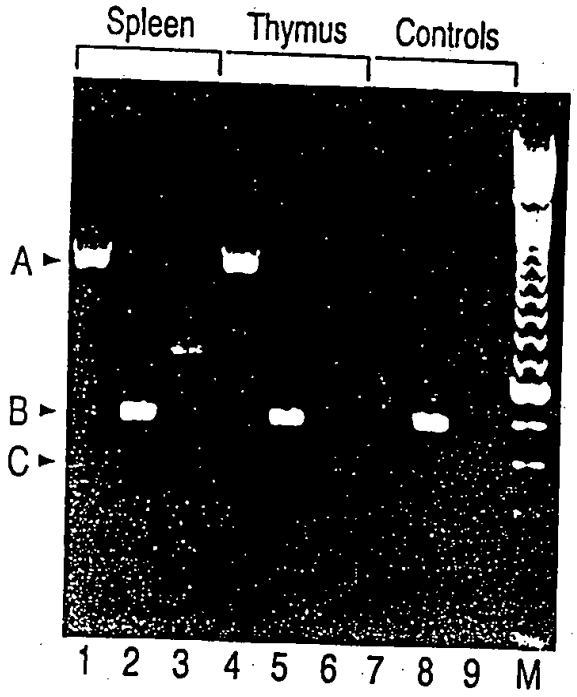

FIGURE 1

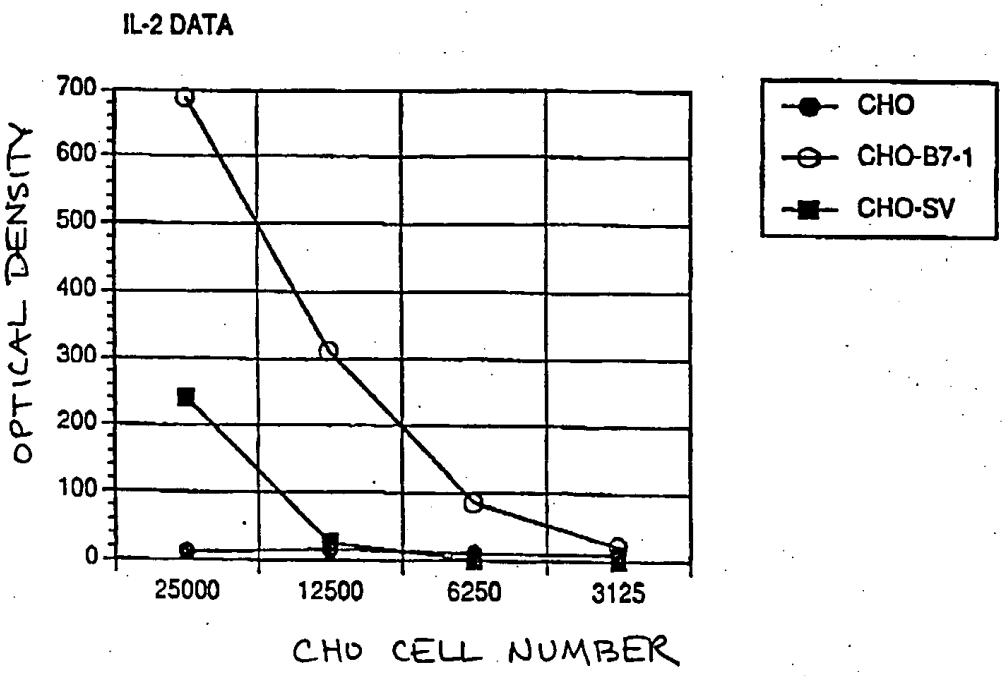
IL-2 DATA
OPTICAL DENSITY
700
600
500
400
300
200
100
0
25000
12500
6250
3125
CHO CELL NUMBER
CHO
CHO-B7-1
CHO-SV

FIGURE 3